RL-TR-94-42
In-House Report
March 1994

STATISTICAL PROPERTIES OF BISTATIC CLUTTER ECHOES

William G. Stevens and Muralidhar Rangaswamy

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

Rome Laboratory
Air Force Materiel Command
Griffiss Air Force Base, New York

This report has been reviewed by the Rome Laboratory Public Affairs Office (PA) and is releasable to the National Technical Information Service (NTIS). At NTIS it will be releasable to the general public, including foreign nations.

RL-TR-94-42 has been reviewed and is approved for publication.

APPROVED: John F. Lennon

JOHN F. LENNON
Acting Chief, Applied Electromagnetics Division
Electromagnetics & Reliability Directorate

FOR THE COMMANDER: Robert V. McGahan

ROBERT V. McGAHAN
Acting Director of Electromagnetics & Reliability

If your address has changed or if you wish to be removed from the Rome Laboratory mailing list, or if the addressee is no longer employed by your organization, please notify RL (ERCE) Hanscom AFB MA 01731. This will assist us in maintaining a current mailing list.

Do not return copies of this report unless contractual obligations or notices on a specific document require that it be returned.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	March 1994	IN-HOUSE, December 1992-December 1993

4. TITLE AND SUBTITLE

STATISTICAL PROPERTIES OF BISTATIC CLUTTER ECHOES

5. FUNDING NUMBERS

PE:	61102F
PROJ NO:	2304
TASK NO:	230414
WORK UNIT:	23041402

6. AUTHOR(S)

William G. Stevens, Muralidhar Rangaswamy

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Rome Laboratory/ERCE
31 Grenier Street
Hanscom AFB, MA 01731-3010

8. PERFORMING ORGANIZATION REPORT NUMBER

RL-TR-94-42

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

In this report we present statistical results of bistatic terrain reflectivity measurements. A new technique for estimating statistical properties from limited data sets was used to examine underlying temporal probability density functions (PDFs) and correlation properties of echoes from several clutter range resolution bins. The case of locally vertically incident-vertically received signal polarization is reported for a single bistatic geometry. The terrain in the experiment consisted primarily of early-development deciduous trees, 10-15 feet high and brush approximately 8 feet high. The measurement system was a high resolution instrumentation radar operating at 3.2 GHz. Results from the estimation algorithm showed that there were small groups of contiguous clutter cells where the members within the group could be described by the same family of PDFs, while other PDF families would best describe neighboring clutter cell groups. Algorithm results for all tested clutter cells showed that the high-resolution, bistatic echoes were non-Rayleigh distributed.

14. SUBJECT TERMS

Bistatic scattering, Terrain scattering measurements, Foliage scattering, Scattering statistics.

15. NUMBER OF PAGES
84

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	SAR

Contents

Illustrations

Statistical Properties of Bistatic Clutter Echoes

1. INTRODUCTION

Investigating the electromagnetic scattering properties of terrain in bistatic geometries through measurement and modeling is necessary to assess the potential of bistatic radars. Clutter models predict that the average power scattered by a rough surface in a given direction other than backscatter differs from the backscatter power level by tens of decibels.[1] Other simulations indicate that the variance of the scattered power also changes for different bistatic configurations.[2] These findings hold for both vertical and horizontal linear incident polarization orientations. These apparent changes in statistical properties lead to questions of underlying statistical distributions of scattered signals for bistatic geometries. In this report we present measurement results of temporal fluctuations in scattered signals for a bistatic scenario. In particular, a new algorithm that provides approximations to underlying statistical distributions of a set of random data is applied to

Received for Publication 11 March 1994

[1] Papa, Robert J., Lennon, John F., and Taylor, Richard L. (1986) The Variation of Bistatic Rough Surface Scattering Cross Section for a Physical Optics Model, *IEEE Trans. Antennas Propagat.*, **AP-3**, (No. 10).

[2] Sharpe, Lisa M. (1991) *Analytical Characterization of Bistatic Scattering From Gaussian Distributed Surfaces*, RL-TR-91-351, AD254253.

uncorrelated clutter measurements of early-growth deciduous foliage. Results are presented for several contiguous resolution cells for the vertical incident - vertical receive polarization case. The novel algorithm uses a comparison of standardized order statistics of the measurement samples with ordered samples drawn from the test distribution. Linked vectors are formed from both measurement and test order statistics and plotted to allow visual assessment of agreement of test distribution with measured data. Results show excellent agreement of the distribution chosen by the algorithm and the histogram of data. The chief advantage of the new algorithm is that it uses very small sample sizes (of order 100).

2. EXPERIMENT DESCRIPTION

Measurements of the temporal fluctuations of a 3.2 GHz signal scattered by a region of early-growth deciduous trees and brush were performed at the Rome Laboratory Ipswich, MA. site. These measurements were conducted in a bistatic geometry with incident angle θ_i of 75 degrees, scattering angle θ_s of 84 degrees, and azimuthal scattering angle ϕ_s of 88.5 degrees, as shown in Figure 1. The azimuthal scattering angle is measured from the forward scatter plane and the origin is the intersection of the boresight of receiving and transmitting antennas on the terrain of interest.

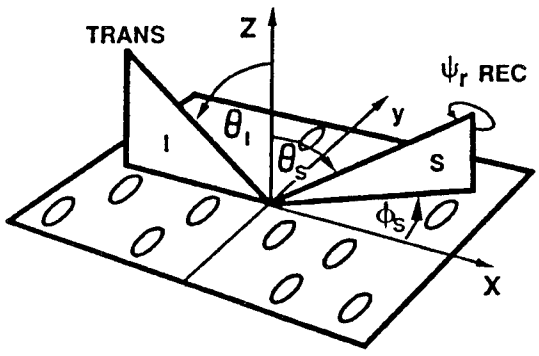


Figure 1. Bistatic Geometry.

The transmitter antenna was elevated 30 feet from the ground and separated from the clutter cell by approximately 140 feet, while the receiver antenna was 523 feet away and 45

feet high. The baseline transmitter and receiver separation was 545 feet. Trees in the clutter cell had an average height of 12 feet. The incident signal was from a dual-linearly-polarized 4-foot diameter parabolic reflector antenna that was electronically switchable between vertical and horizontal linear polarization. The receive antenna was a 6 foot diameter dual-linearly-polarized parabolic reflector. The measured cross-polarization isolation at boresight was -25 dB for both the transmit and receive antennas.

The measurements were conducted with a high resolution instrumentation radar system. A 1023 bit Binary Phase Shift Keying code modulated the 3.2 GHz continuous wave signal to allow range resolving capability without requiring the higher peak power of a conventional pulsed radar. Each bit in the code was 5 nanoseconds long, allowing clutter echoes separated by as little as 4.9 feet to be resolved.

The receiver used a correlation detector to extract the clutter echo amplitude and phase information from the pseudonoise waveform. With this technique a waveform with a code pattern identical to that which was transmitted is generated in the receiver and cross-correlated with the signal reflected from each resolution cell in the clutter. The correlation detection process was repeated for each of the 1023 resolution cells, with 200 milliseconds - equivalent to the pulse repetition interval - required for demodulating the echoes from all of the cells.

3. ALGORITHM DESCRIPTION

We briefly outline the algorithm used for the statistical analysis of radar clutter data in this section. The reader is referred to Appendix A for mathematical details and to the references for a thorough description.[3,4] Statistical characterization of radar clutter is important from both analysis and system design standpoints. From an analysis point of view, we are interested in determining the physics of the scattering mechanism that gives rise to the clutter. From a system design point of view, we are interested in determining the optimal radar signal processor that enables target detection in a given clutter environment.[5] Statistical characterization of radar clutter enables us to achieve both of these objectives.

More precisely, we are interested in determining the underlying probability density function (PDF) of a set of radar clutter data. In general, this problem does not have a unique solution. Currently available approaches such as the Kolmogorov-Smirnov and chi-square tests address the problem of goodness-of-fit to a set of random data. In particular, they

---

[3] Shah, Rajiv R. (1993) *A New Technique for Distribution Approximation of Radar Data*, M.S. Thesis, Syracuse University.

[4] Slaski, Lisa, and Rangaswamy, Muralidhar, (RL Report in Preparation) *An Introduction to Dr. Ozturk's Algorithm for PDF Approximation.*

[5] Rangaswamy, M., Chakravarthi, P., Weiner, D.D., Cai, L., Wang, H., and Ozturk, A. (1993) *Signal Detection in Correlated Gaussian and Non-Gaussian Radar Clutter*, RL-TR-93-79, AD267453.

provide an answer to the question "is a set of data statistically consistent with a specified PDF?" However, if the answer to the above question is negative, these tests do not provide a PDF that approximates the PDF of the set of data. Furthermore, these tests require a large number of samples for satisfactory performance.

The algorithm developed in Appendix A is used to address the problem of statistical characterization of radar clutter measurements made using the approach of Section 2. This algorithm has two modes of operation. In the first mode, the algorithm performs a goodness-of-fit test. Specifically, the test determines, to a desired confidence level, whether a set of data is statistically consistent with a specified PDF. In the second mode of operation, the algorithm approximates the PDF of a set of data. In particular, by analyzing the data and without any *a priori* knowledge, the algorithm identifies, from a stored library of PDFs, the particular density function that best approximates the data. Estimates of the scale, location, and shape parameters of the approximating PDF are provided by the algorithm. Both modes of operation of the algorithm are graphical and provide a visual representation of the goodness-of-fit and distribution approximation techniques. Of particular note is the observation that the algorithm works well with as few as 100 samples.

The algorithm is based on the assumption that we are dealing with independent, identically distributed random variables. Currently available tests for statistical independence can be applied only to Gaussian random variables. However, it is likely that the data encountered in this analysis are non-Gaussian. Therefore, statistical independence of the data is not guaranteed. On the other hand, it is possible to determine the correlation properties and spectral characteristics (using an FFT) of the set of data by estimating the correlation function and the power spectral density. These estimates enable us to determine the correlation time of the clutter process and allow us to use uncorrelated data samples for the algorithm. The results of the algorithm are independently verified by the use of a histogram on the set of uncorrelated data.

4. RESULTS

The new algorithm of Appendix A was applied to 100 uncorrelated data points from each of nine range bins to perform the test of goodness-of-fit to the Gaussian distribution and, if the data are rejected as Gaussian, to estimate the underlying distribution. The algorithm provides 27 different approximating PDFs to the data set.

Data from range bins 8 through 16 were chosen due to constraints imposed by the bistatic geometry and antenna patterns. Results for each range bin are presented in groups of six separate figures. The first figure in the group illustrates the time sequence of the clutter echoes represented in amplitude-phase, and real-imaginary component forms. These are the raw clutter returns for the 1000 consecutive data frames collected over a 200 second period, and are not necessarily uncorrelated measurements. The second figure depicts the unbiased autocorrelation sequence estimate of the complex clutter echo versus lag number. The results of this sequence are used to determine the decorrelation time of the data from a

given clutter cell. This decorrelation time is determined by counting the number of time lags that occur for the autocorrelation sequence to decrease from 1.0 to 0.1.

The third figure presents temporal histograms of the magnitude, phase, in-phase component, and quadrature component of the uncorrelated clutter returns. This allows visual assessment of characteristics such as the uniformity of the distribution of phases.

The fourth figure in the group illustrates the graphical technique for determining the goodness-of-fit of the data to the null hypothesis distribution, which in this case is the Gaussian distribution. The goodness-of-fit chart is constructed by arranging the vectors derived from the sample in order of their size, and plotting them to make a trajectory. Another set of vectors, also arranged in order of their size, is plotted for the Gaussian distribution assumed as the null hypothesis. Confidence contours are plotted around the end point of the null hypothesis trajectory. Terminal points of the data trajectory falling into the area contained by the outermost ellipse correspond to a probability of 0.01 that the data are not represented by the null hypothesis. Terminal data points contained by the middle ellipse indicate that with probability 0.05, the data are not described by the null hypothesis. Termination of the data trajectory within the innermost ellipse, corresponds to a probability of 0.1 that the data is not described by the null hypothesis. If a terminal sample point falls inside the appropriate ellipse for the confidence level desired, the data are considered consistent with the Gaussian distribution, with a confidence level of [1 minus (the probability for that ellipse)]. On these figures, the confidence levels would be 0.99, 0.95, and 0.9 for the outer, middle, and inner ellipses, respectively. If the terminal point falls outside the ellipse, the null (Gaussian) hypothesis is rejected, with a significance equal to probability represented by the ellipse. Although the terminal point of the linked vector is plotted in the fifth figure of each group, showing its location on the PDF approximation chart, the shape of the trajectory shown in Figure 4 is also used to determine whether the data are statistically consistent with the null hypothesis. A trajectory for data that are consistent with the null hypothesis should not get farther from the null hypothesis trajectory than the distance between the terminal points of the sample and null hypothesis trajectories.

The fifth figure of each group is the PDF approximation chart. Each curve represents the linked-vector endpoint trajectory for one probability density function as the shape parameter is varied. For PDFs with no shape parameter, the linked vector trajectory appears as a single point on the chart. For PDFs with more than one shape parameter, a family of curves are generated on the chart. For example, the beta distribution has two shape parameters. In this case, the family of curves is obtained by fixing the first shape parameter at its minimum value and varying the second. A second value is then assigned to the first shape parameter, and the second shape parameter is again varied. As more values are assigned to the first shape parameter, a family of curves is generated. The last curve in the family is generated by assigning the maximum value to the first shape parameter and varying the second shape parameter. Thus, a family of curves corresponding to all possible values of the shape parameters of the beta distribution is shown in the chart. The large 'X' on the chart is the linked-vector endpoint of the sample data set. The last figure of the group

overlays the algorithm's first and last choices for best approximating PDF onto a histogram of the sample data, based on parameter estimates that are also provided by the algorithm.

Figures 2 through 7 present results of the echo from delay resolution cell 8. In Figure 2, the magnitude is seen to vary by a factor of approximately 6, and the phase varies over the entire range. The correlation sequence of Figure 3 shows the time of decorrelation to 0.1 is approximately 4.5 lags. The histograms of Figure 4 show that the phase is nearly uniformly distributed and the magnitude distribution exhibits a non-Rayleigh trend, as has been observed in previous measurements of high resolution clutter. The goodness-of-fit chart in Figure 5 shows that the data fall within the confidence contour of 0.1 probability of not satisfying the null hypothesis. It can be inferred from this chart that a non-Gaussian distribution probability best represents the amplitude statistics. The location of the 'X' on the PDF approximation chart of Figure 6 shows that the amplitude fluctuations most closely follow lognormal statistics, as determined by the algorithm.

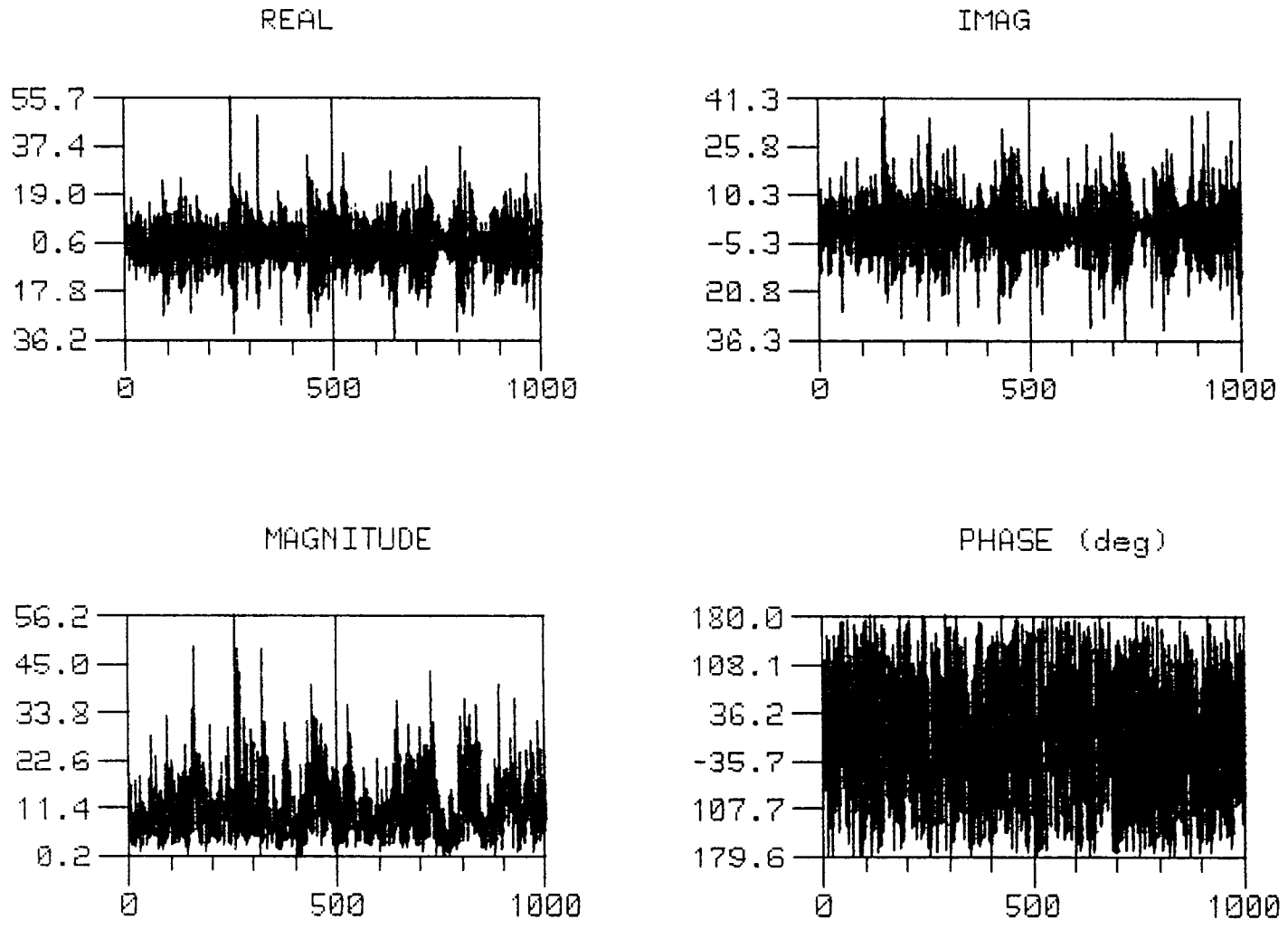


Figure 2. Raw Data of Bin 8.

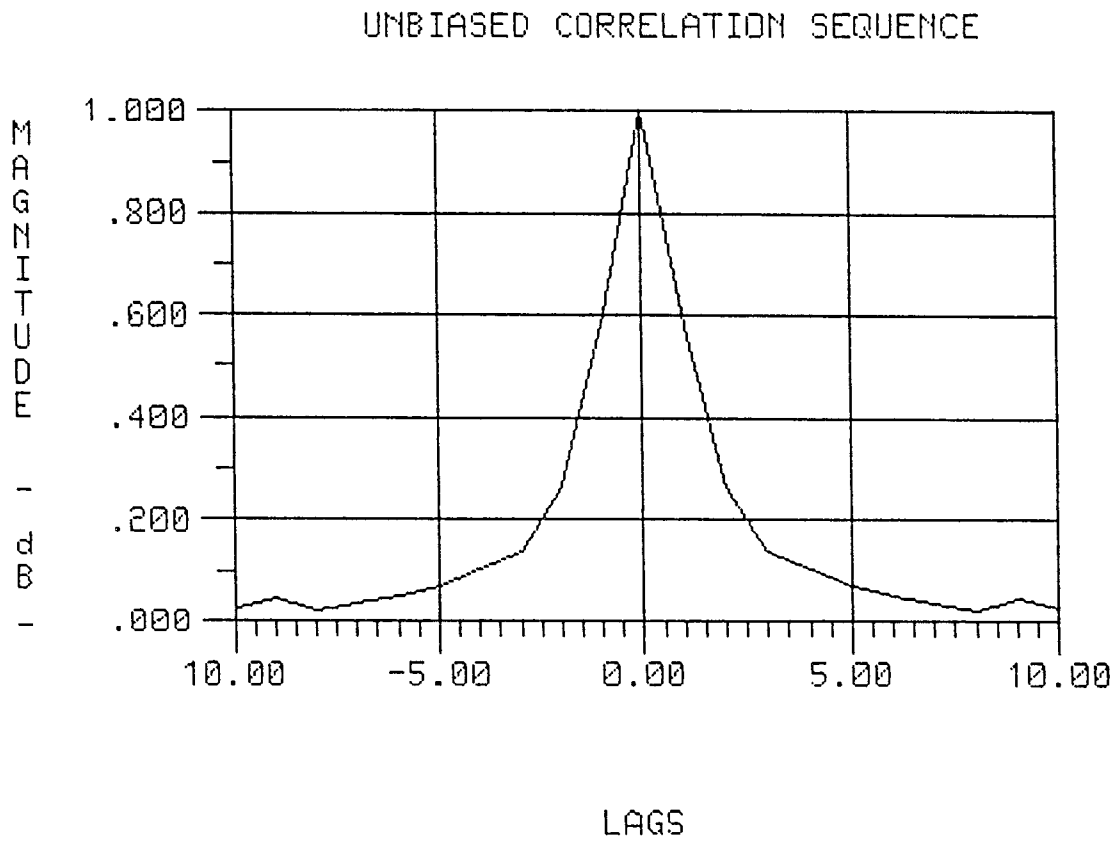


Figure 3. Correlation Sequence Estimate of Bin 8 Data.

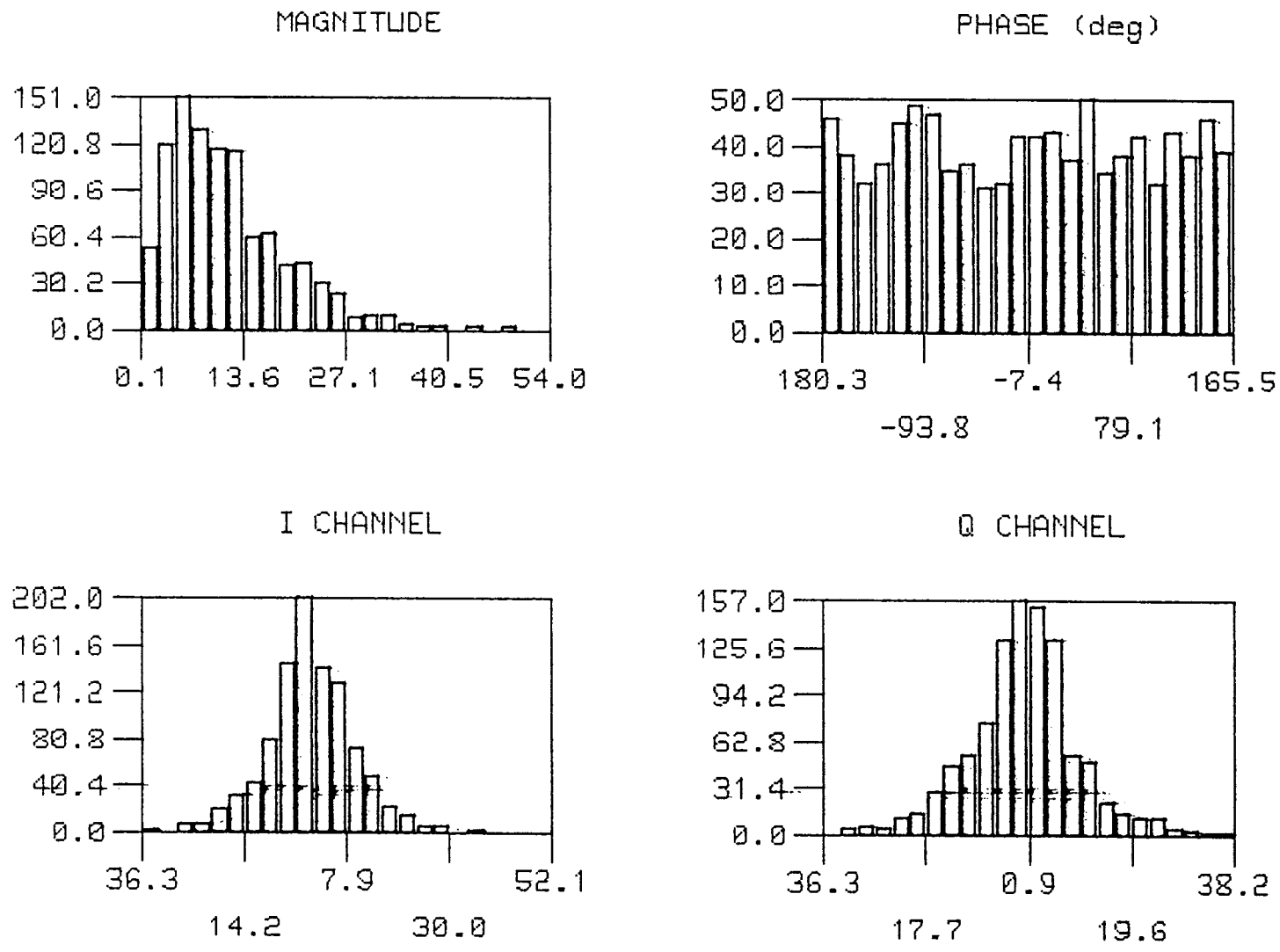


Figure 4. Histograms of Bin 8 Raw Data.

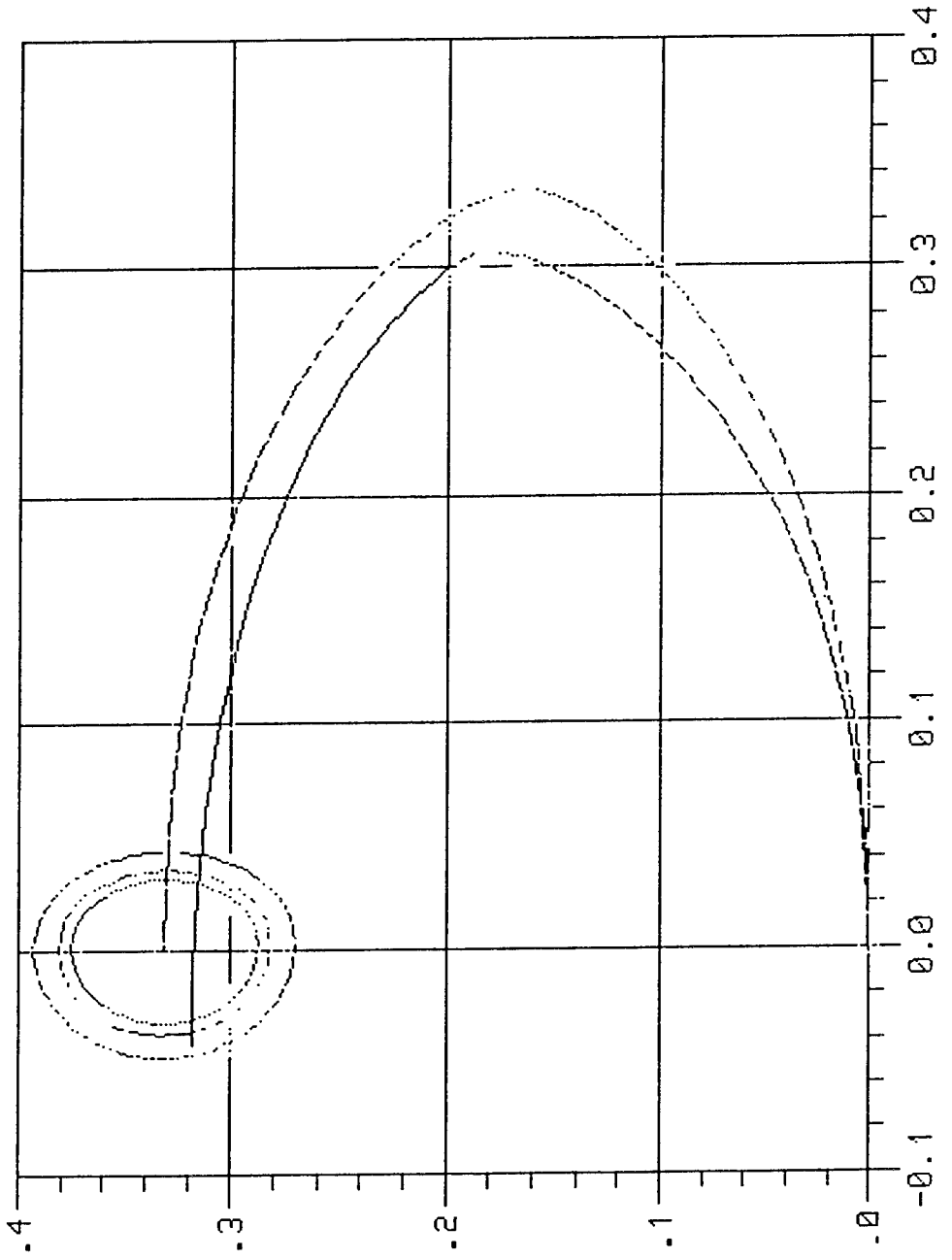
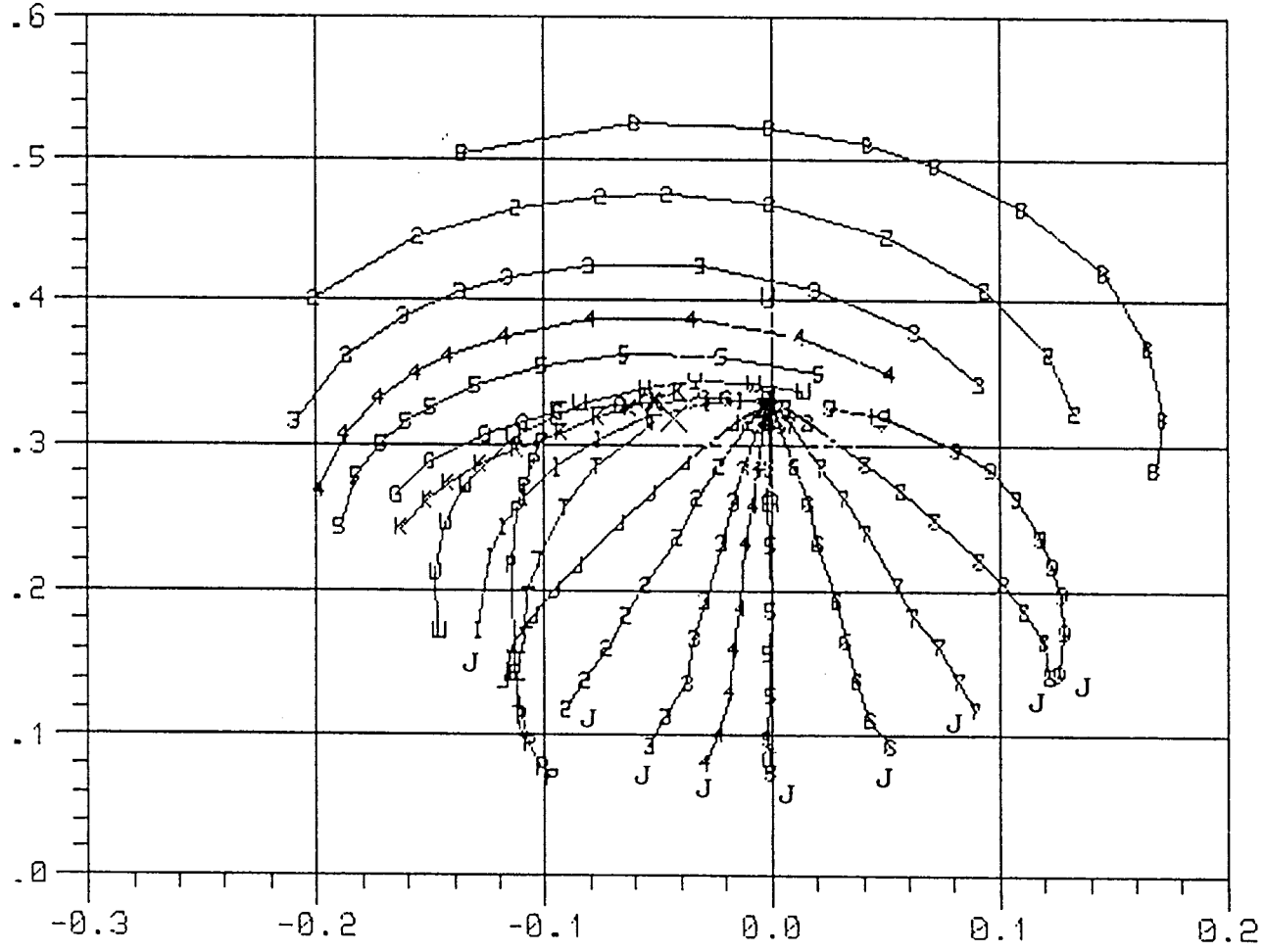


Figure 5. Goodness-Of-Fit Plot for Bin 8.



Nomenclature Indicates Type of Distribution

B, 2, 3, 4, and 5 – Beta Distribution with different values for the first shape parameter
G – Gamma Distribution
K – K Distribution
W – Weibull Distribution
P – Pareto Distribution
T – Extreme Value Distribution
L – LogNormal Distribution
U – Uniform Distribution
E – Exponential Distribution
J – SU-Johnson System (family of 9 curves)

Figure 6. PDF Approximation Chart for Bin 8.

As a final visual inspection, the best and worst candidate PDFs with algorithm-chosen parameters are overlaid onto a histogram of the data in Figure 7. Even with the limited number of samples, the best candidate PDF is a very good approximation to the data.

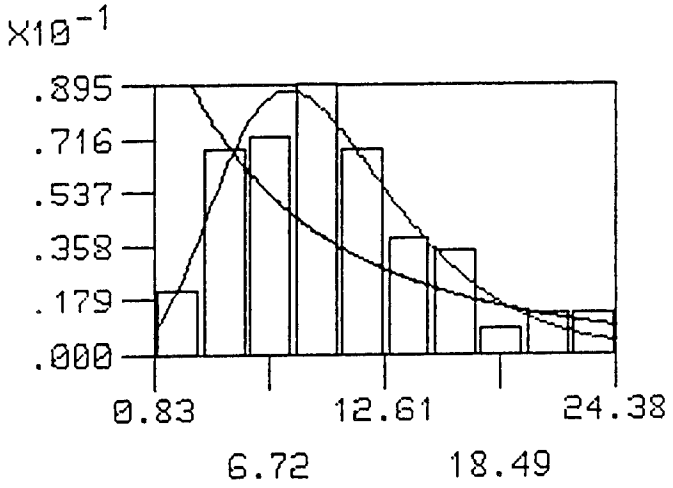


Figure 7. Overlay of Best/Worst PDF Approximations for Bin 8.

Figures 8 through 13 contain various results of the echoes from resolution cell 9. In Figure 8, the magnitudes are seen to vary over the same amplitude range as those of cell 8, but the first 400 samples of cell 9 have a consistently lower magnitude. The phase characteristics of cell 9 also show slower fluctuations than those of cell 8. The correlation time as determined from the correlation sequence of Figure 9 is about 5 lags. This is similar to cell 8. The temporal histograms of Figure 10 show characteristics similar to those of cell 8; namely, near uniformly-distributed phase statistics and non-Gaussian temporal fluctuations. As seen in Figure 11, the amplitude fluctuations are not likely to obey Gaussian statistics. This is verified by the PDF approximation chart of Figure12, where the candidate PDF recommended by the algorithm is Weibull. A visual check of Figure 13 reinforces the acceptance of the Weibull distribution as the closest fit to the sample data.

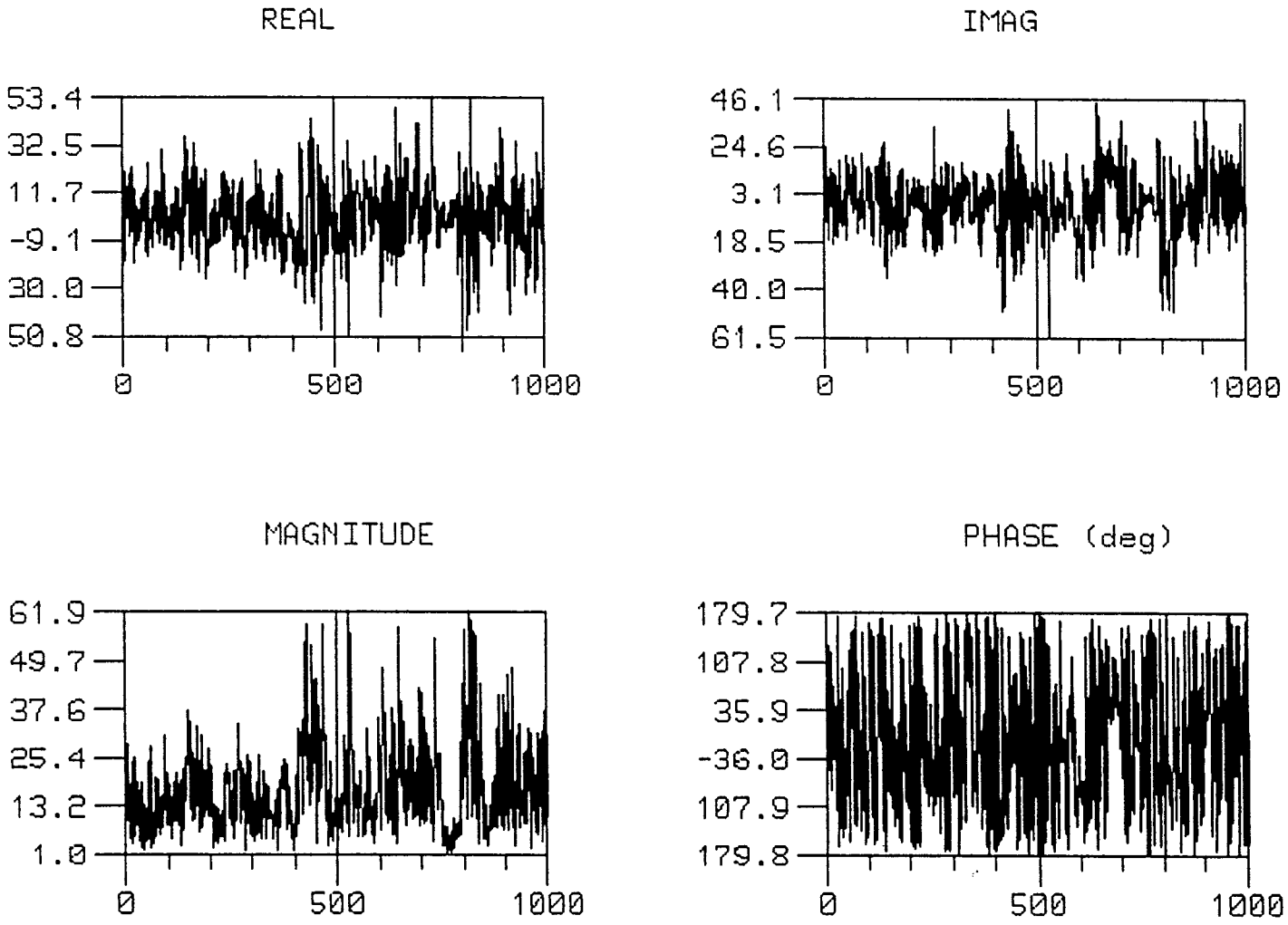


Figure 8. Raw Data of Bin 9.

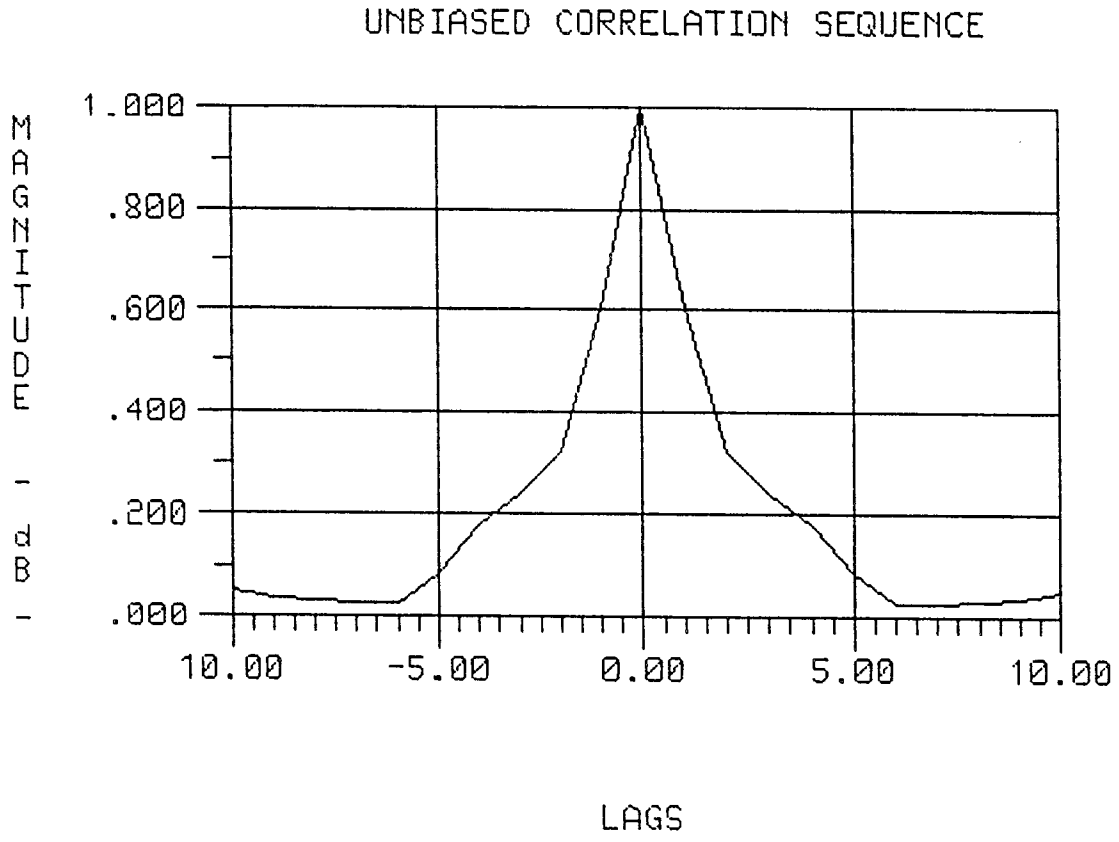


Figure 9. Correlation Sequence Estimate of Bin 9 Data.

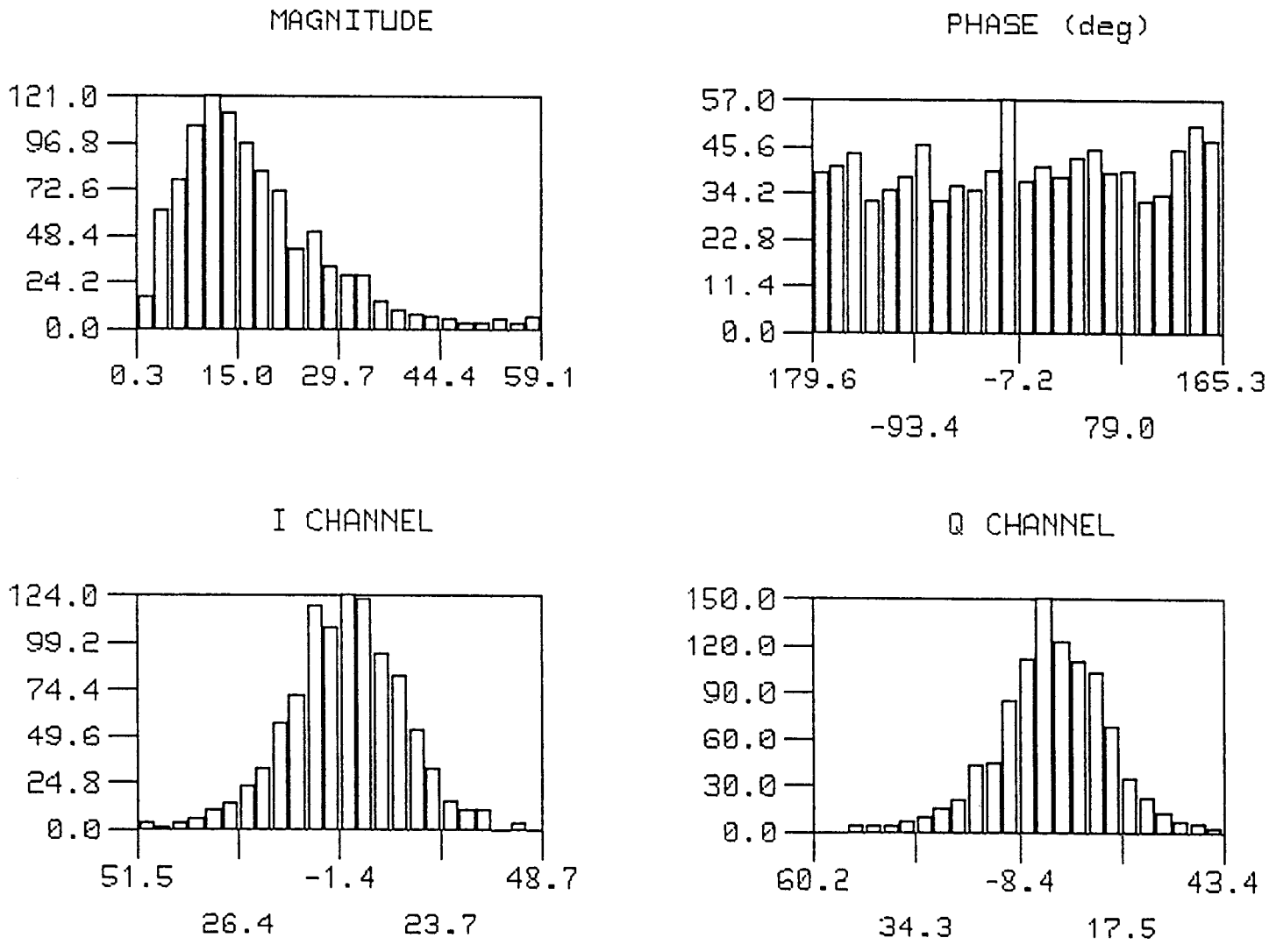


Figure 10. Histograms of Bin 9 Raw Data.

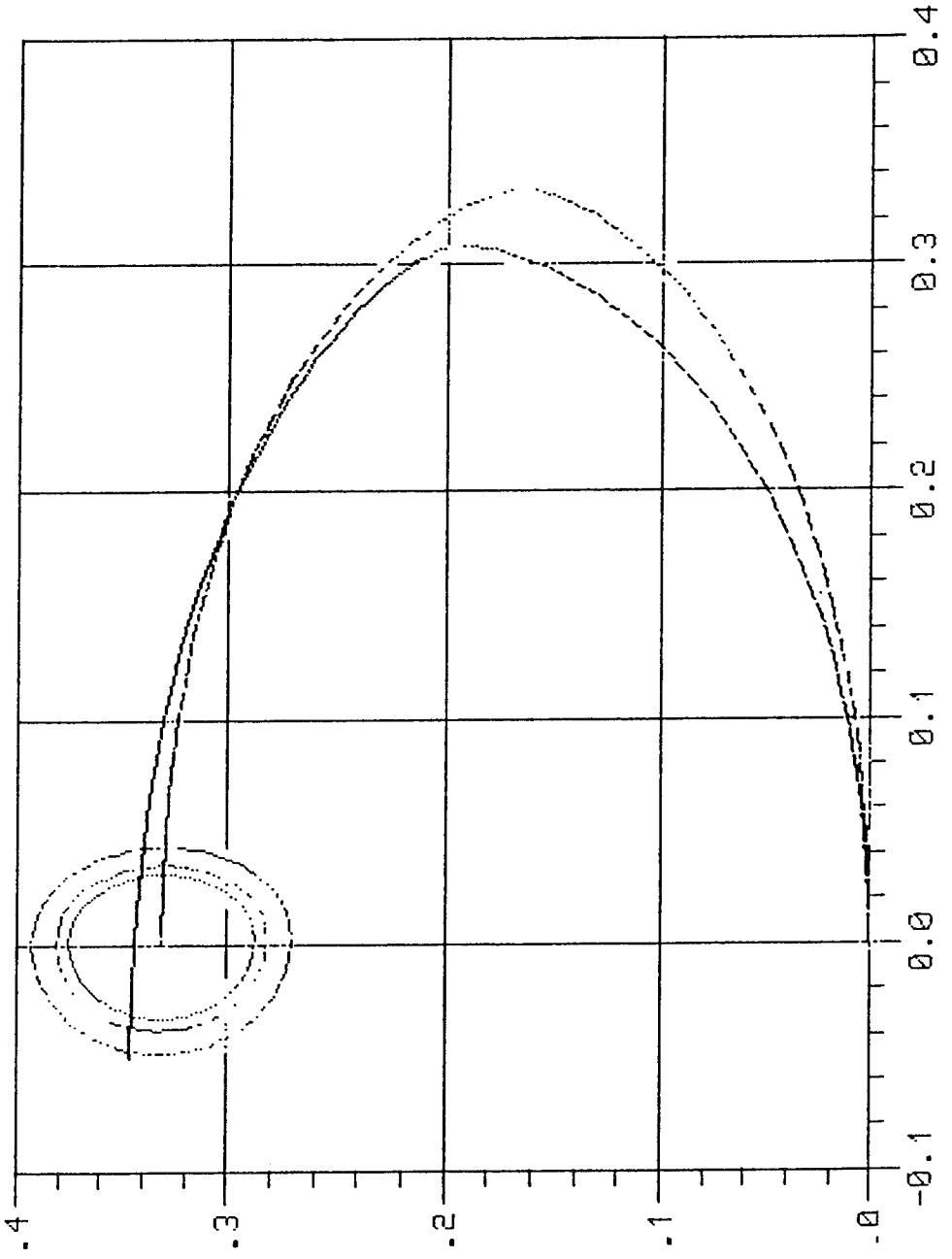
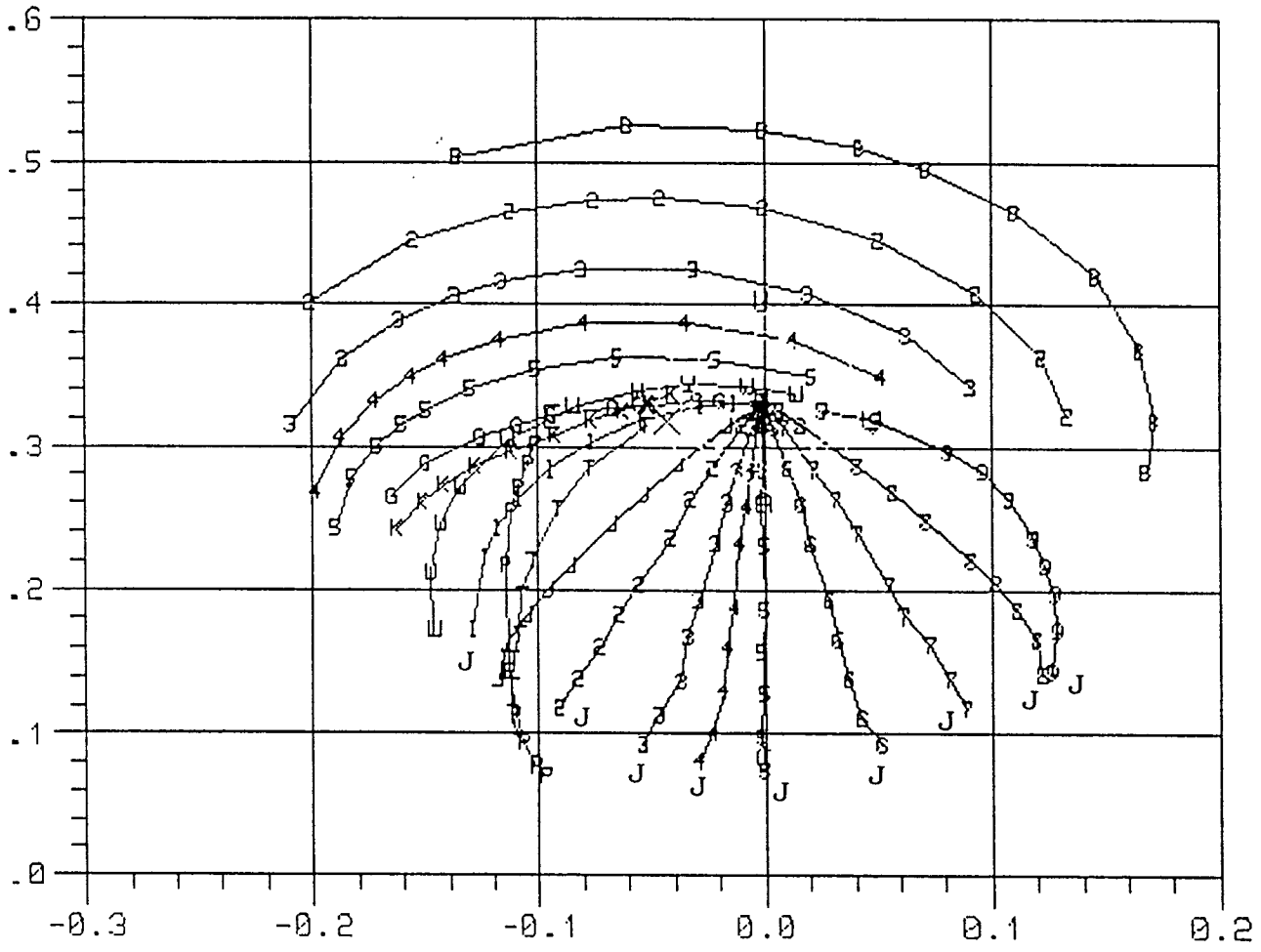


Figure 11. Goodness-Of-Fit Plot for Bin 9.



Nomenclature Indicates Type of Distribution

B, 2, 3, 4, and 5 – Beta Distribution with different values for the first shape parameter
G – Gamma Distribution
K – K Distribution
W – Weibull Distribution
P – Pareto Distribution
T – Extreme Value Distribution
L – LogNormal Distribution
U – Uniform Distribution
E – Exponential Distribution
J – SU-Johnson System (family of 9 curves)

Figure 12. PDF Approximation Chart for Bin 9.

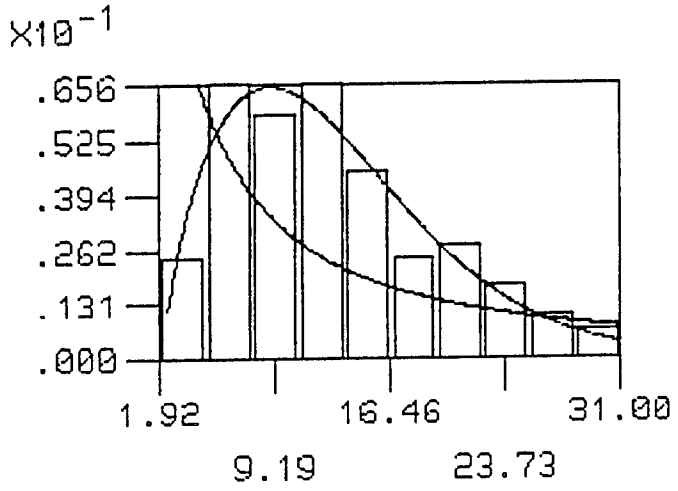


Figure 13. Overlay of Best/Worst PDF Approximations for Bin 9.

Results from resolution cell 10 are presented in Figures 14 through 19. The raw data of Figure 14 show the same range of amplitude fluctuations as cells 8 and 9, but with fewer occurrences of the high amplitude echoes. The temporal phase variation more closely resembles that of resolution cell 9 - the slow, patterned variation. The correlation sequence plot of Figure 15 is similar to that of cell 8, with the same decorrelation time of 4.5 lags. The temporal phase histogram of Figure 16 is nearly uniform, as was the previous cases. The histogram of amplitudes in this figure shows a trend similar to previous cells for the lower amplitude region, but there is less of the larger magnitude tail structure. In the goodness-of-fit plot of Figure 17, the null hypothesis PDF was Weibull instead of Gaussian, as in the goodness-of-fit tests for previous clutter cells. The figure shows that the terminal linked vector of the sample data is very close to that of the approximating PDF and is well within the 0.01 confidence contour of the null hypothesis. The location of the sample data linked vector terminal point (the 'X') on the PDF approximation chart in Figure 18 is closest to the Weibull trajectory. This corroborates the results of the goodness-of-fit test, where the sample data were verified to be statistically consistent with the Weibull distribution. Comparison of best and worst candidate PDFs with measured data is shown in Figure 19. Again, the PDF with family and shape chosen by the algorithm match well with the limited data set.

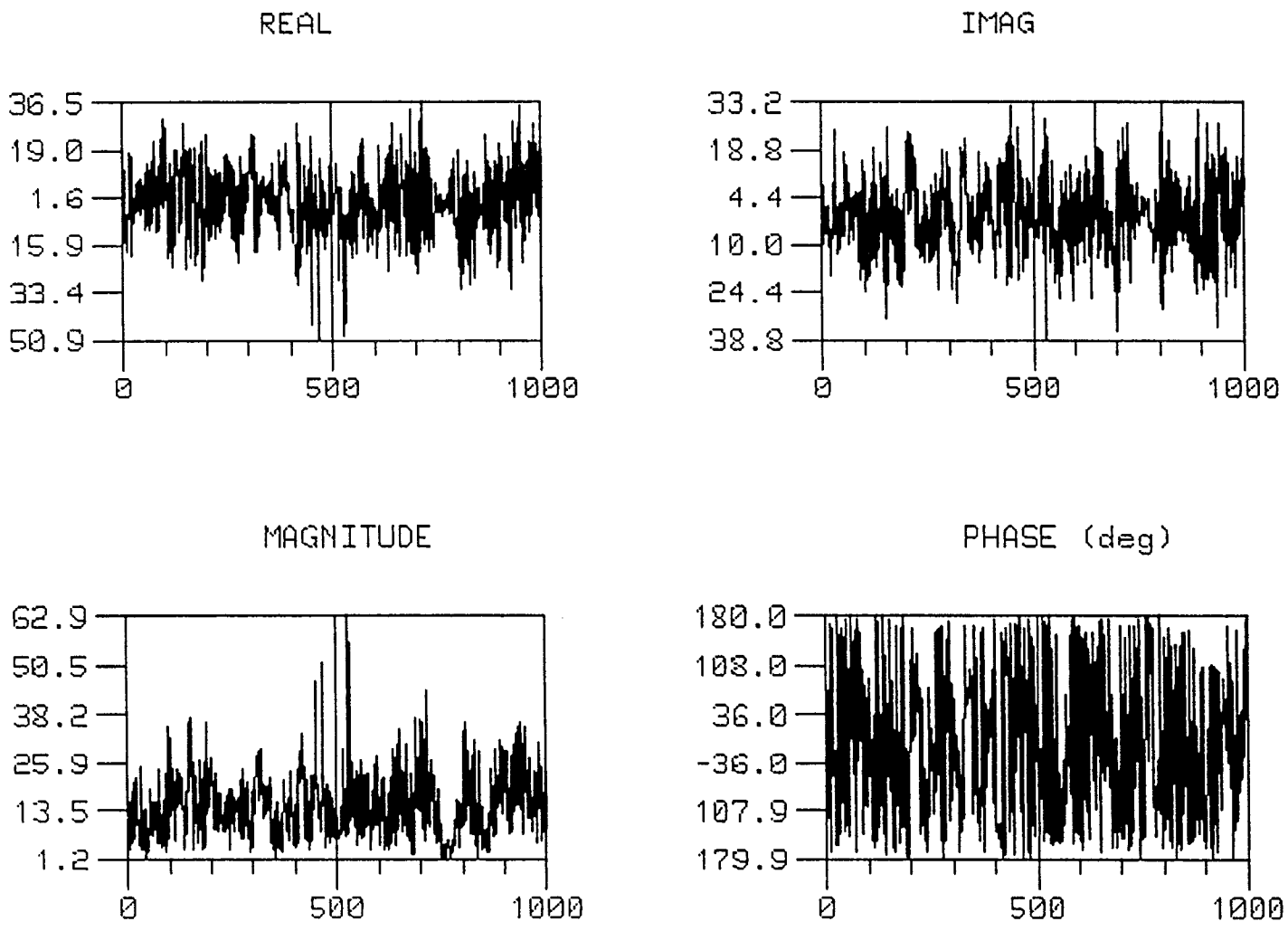


Figure 14. Raw Data of Bin 10.

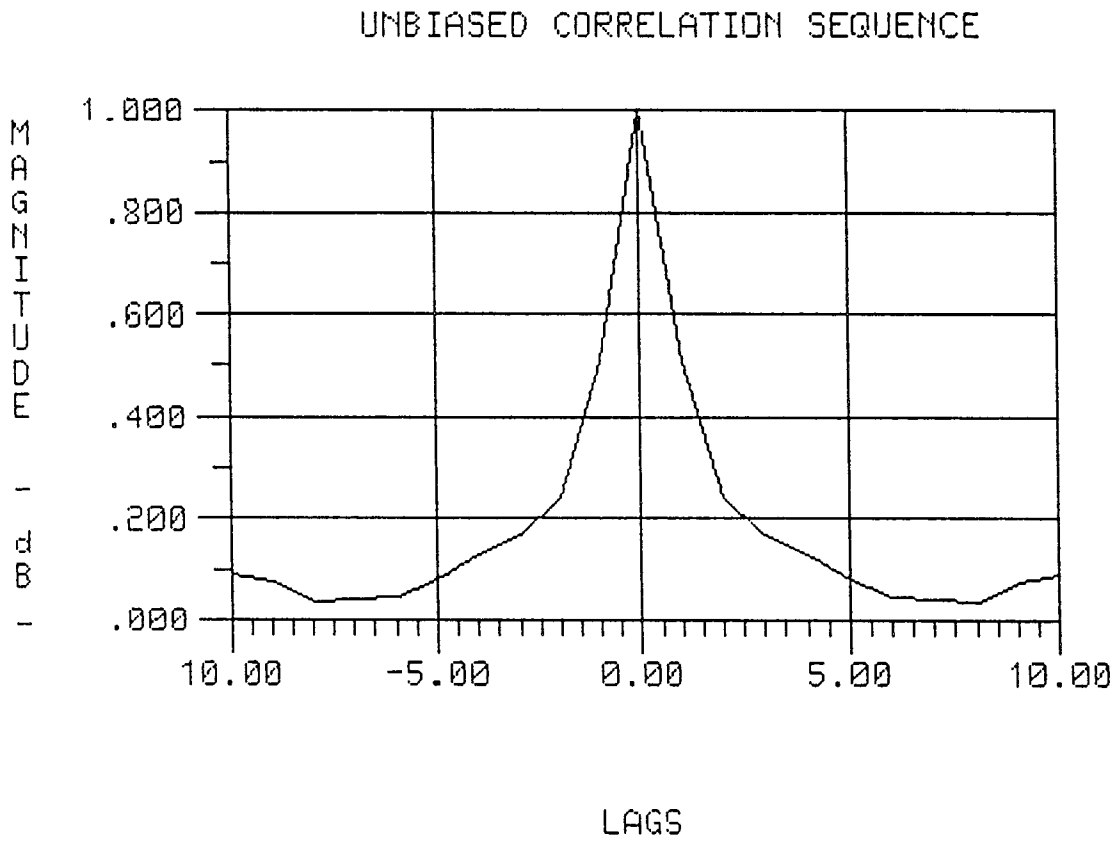


Figure 15. Correlation Sequence Estimate of Bin 10 Data.

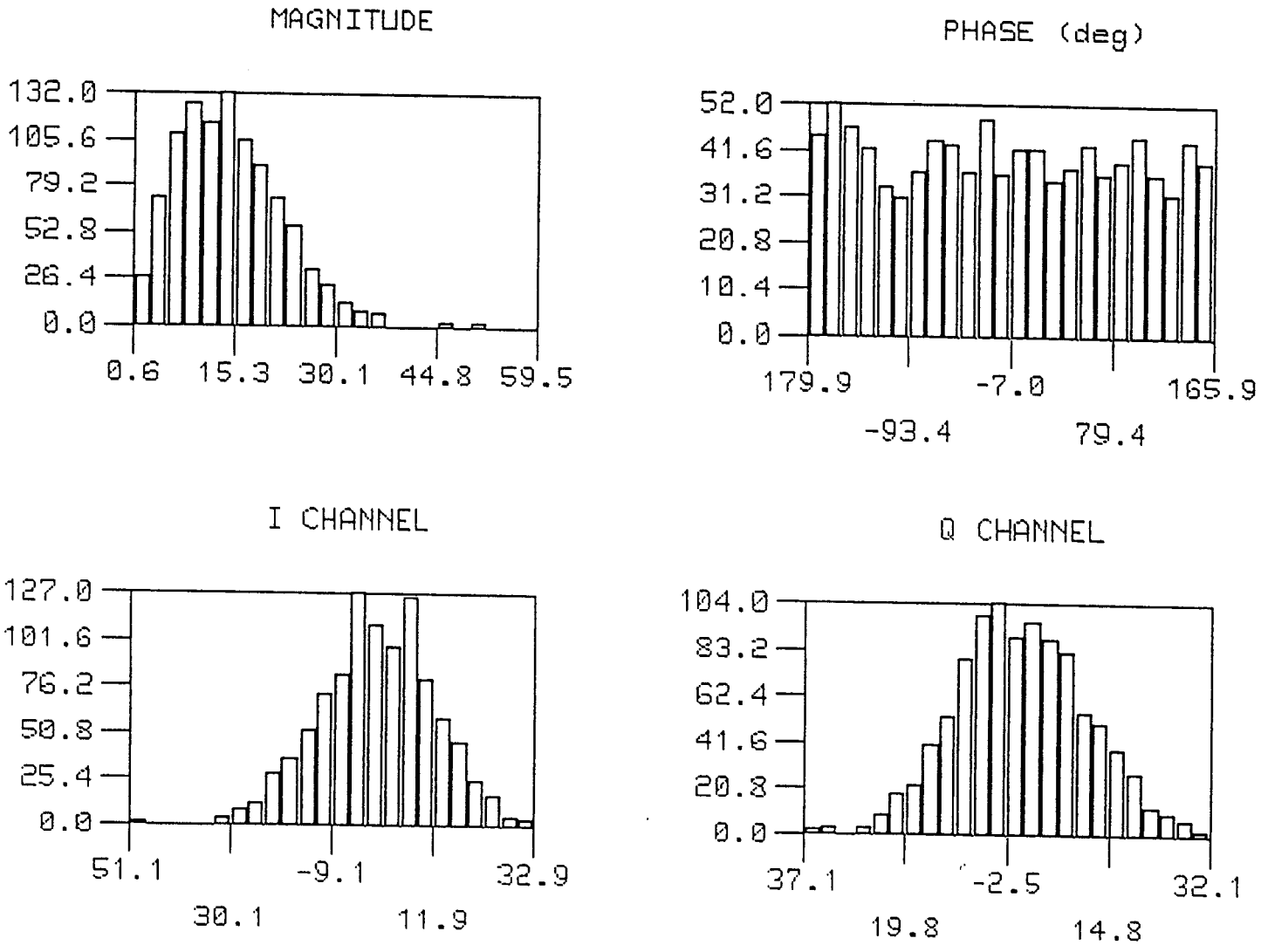


Figure 16. Histograms of Bin 10 Raw Data.

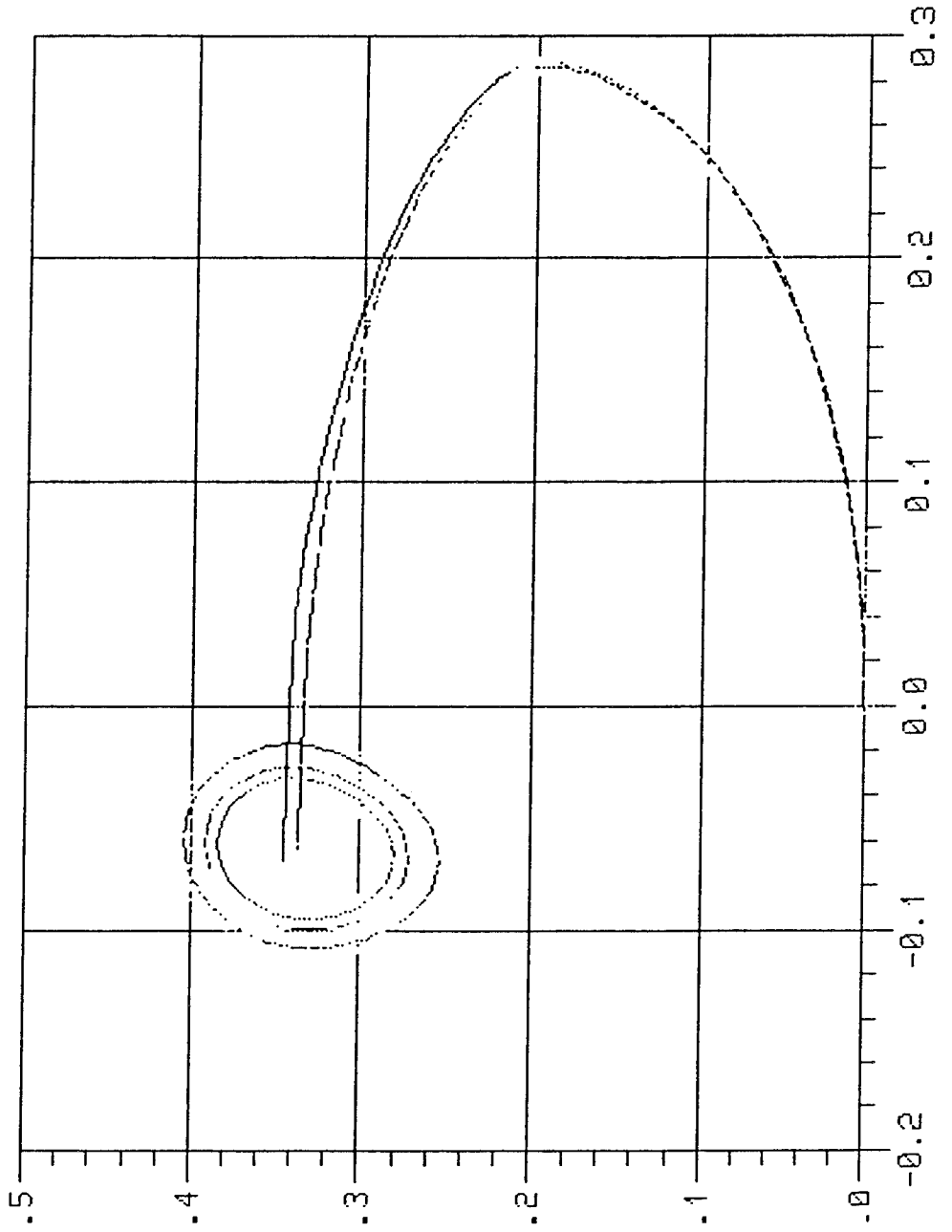
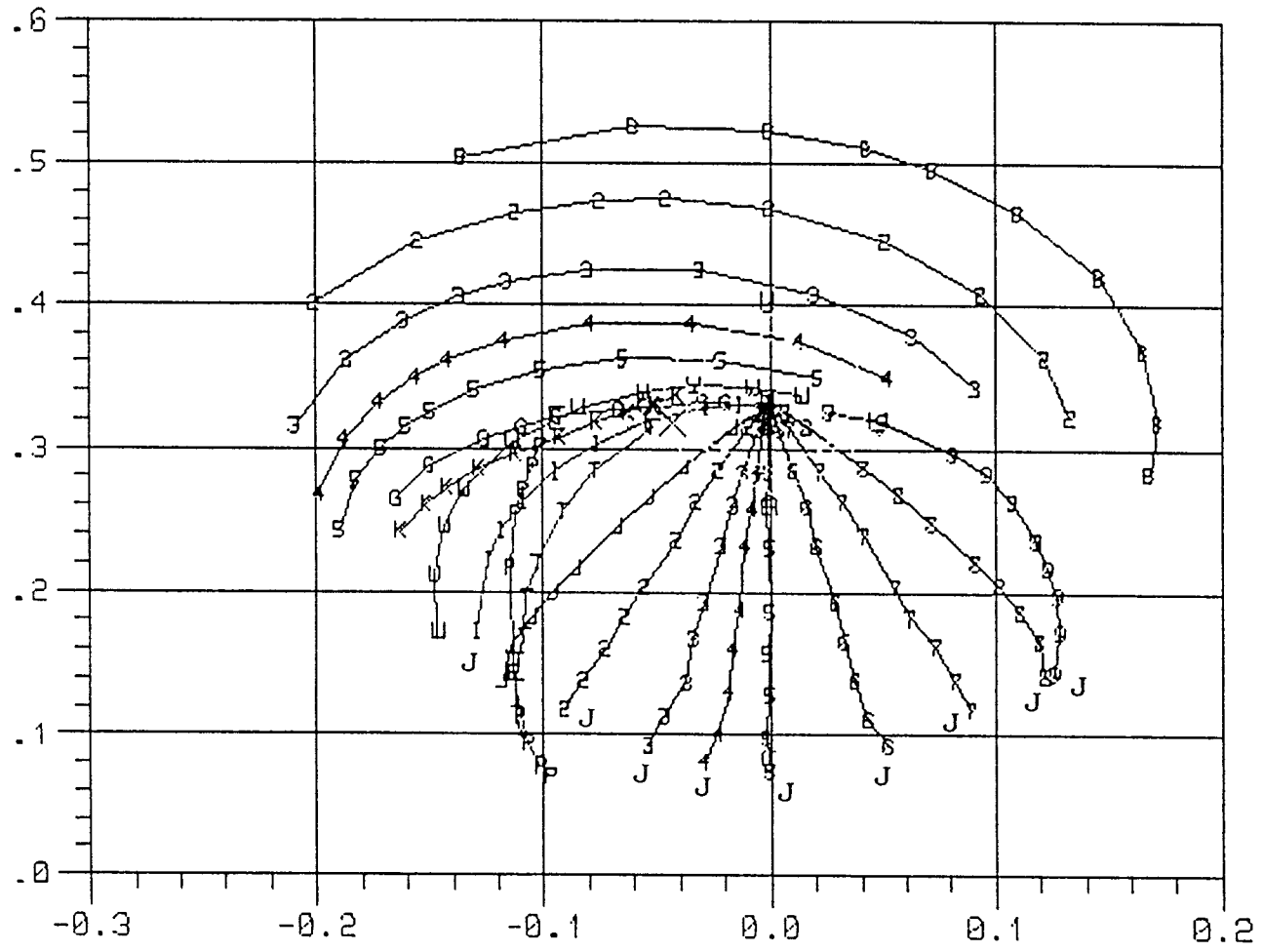


Figure 17. Goodness-Of-Fit Plot for Bin 10.



Nomenclature Indicates Type of Distribution

B, 2, 3, 4, and 5 – Beta Distribution with different values for the first shape parameter
G – Gamma Distribution
K – K Distribution
W – Weibull Distribution
P – Pareto Distribution
T – Extreme Value Distribution
L – LogNormal Distribution
U – Uniform Distribution
E – Exponential Distribution
J – SU-Johnson System (family of 9 curves)

Figure 18. PDF Approximation Chart for Bin 10.

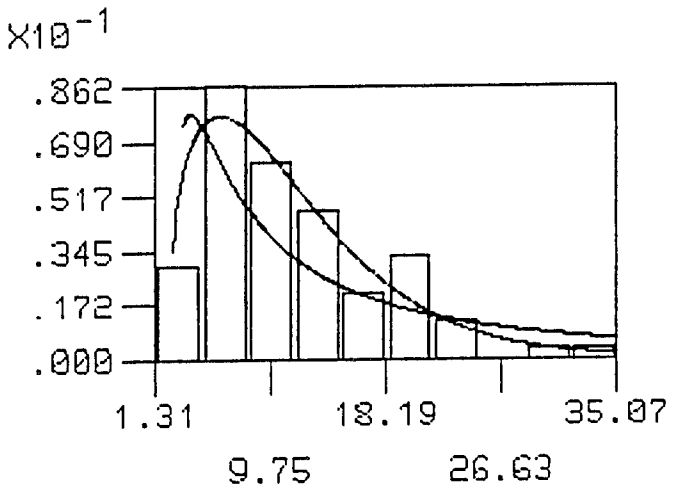


Figure 19. Overlay of Best/Worst PDF Approximations for Bin 10.

Figures 20 through 25 illustrate the sequence of analysis results on clutter resolution cell 11. The amplitude peaks shown in Figure 20, with a maximum of approximately 39 volts^2, are generally lower than those of the previous clutter cells that have maximum values of around 56 volts^2. The phase appears to vary more quickly with time than the previous few resolution cells. The correlation sequence shown in Figure 21 has a trend similar to that of previous cells, but with a longer tail. This produces a time of decorrelation to 0.1 of about 7.5 lags. This is 2.5 lags, or 0.5 seconds, greater than the decorrelation time of resolution cell 9. The temporal phase histogram in Figure 22 shows a nearly uniform distribution, as was seen in previous clutter cells. The amplitude distribution also appears similar to those of previous cells, but with a lower magnitude tail. Visual appearance of a histogram can be inaccurately interpreted, as was seen for previous cells where the histograms had a similar structure but were determined to be from different classes of probability density functions. In the goodness-of-fit test of Figure 23, the sample data are clearly inconsistent with the Gaussian null hypothesis. From the PDF approximation chart in Figure 24, the sample data are determined to be from the Weibull distribution. The sample data linked vector terminal point appears to be nearly equidistant from a beta and the Weibull trajectories, but an actual distance calculation between the data point and the two curves shows the data to be closer to the Weibull distribution. The overlay onto the data histogram of best and worst candidate PDF approximations, shown in Figure 25, again illustrates a successful decision by the algorithm.

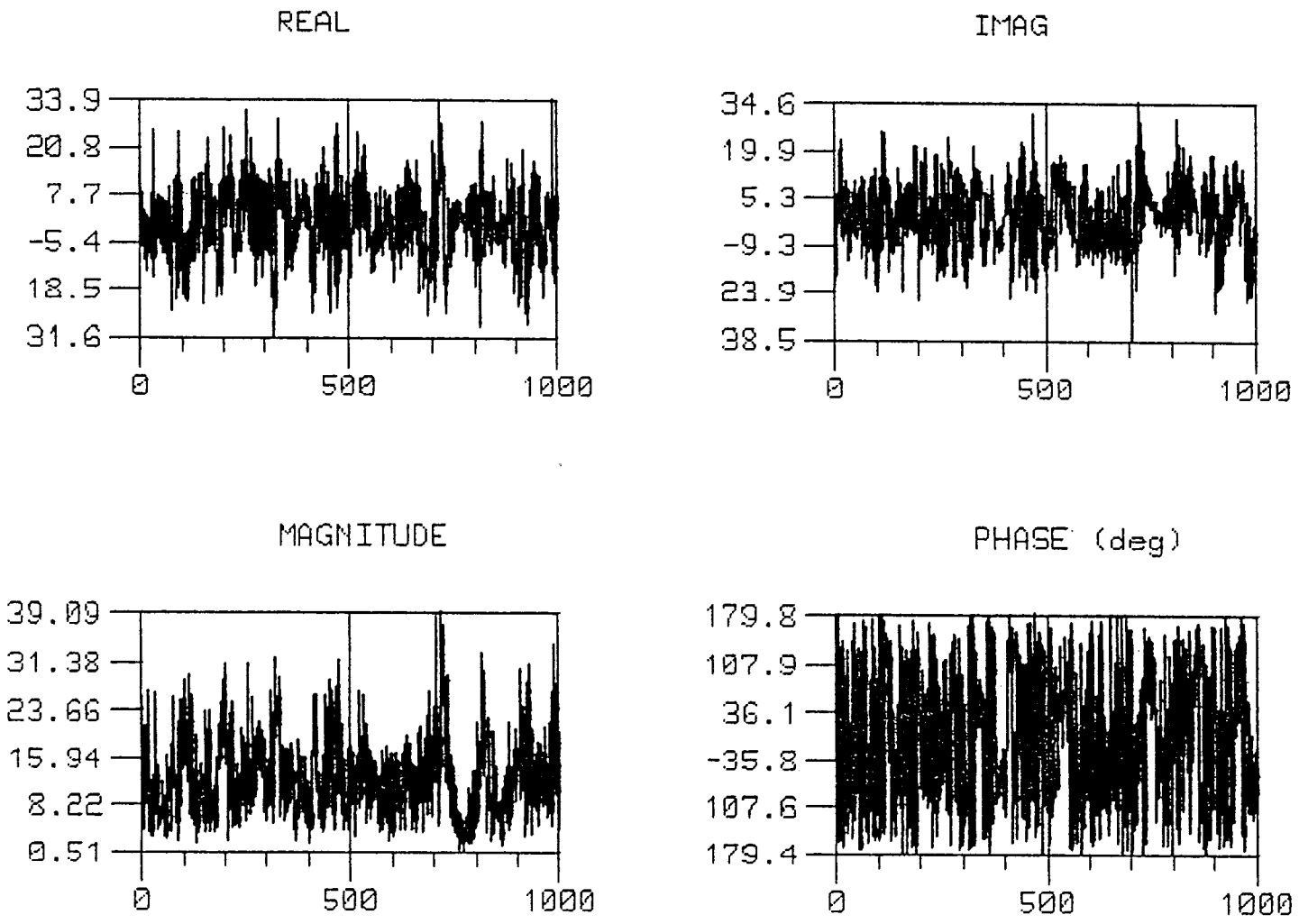


Figure 20. Raw Data of Bin 11.

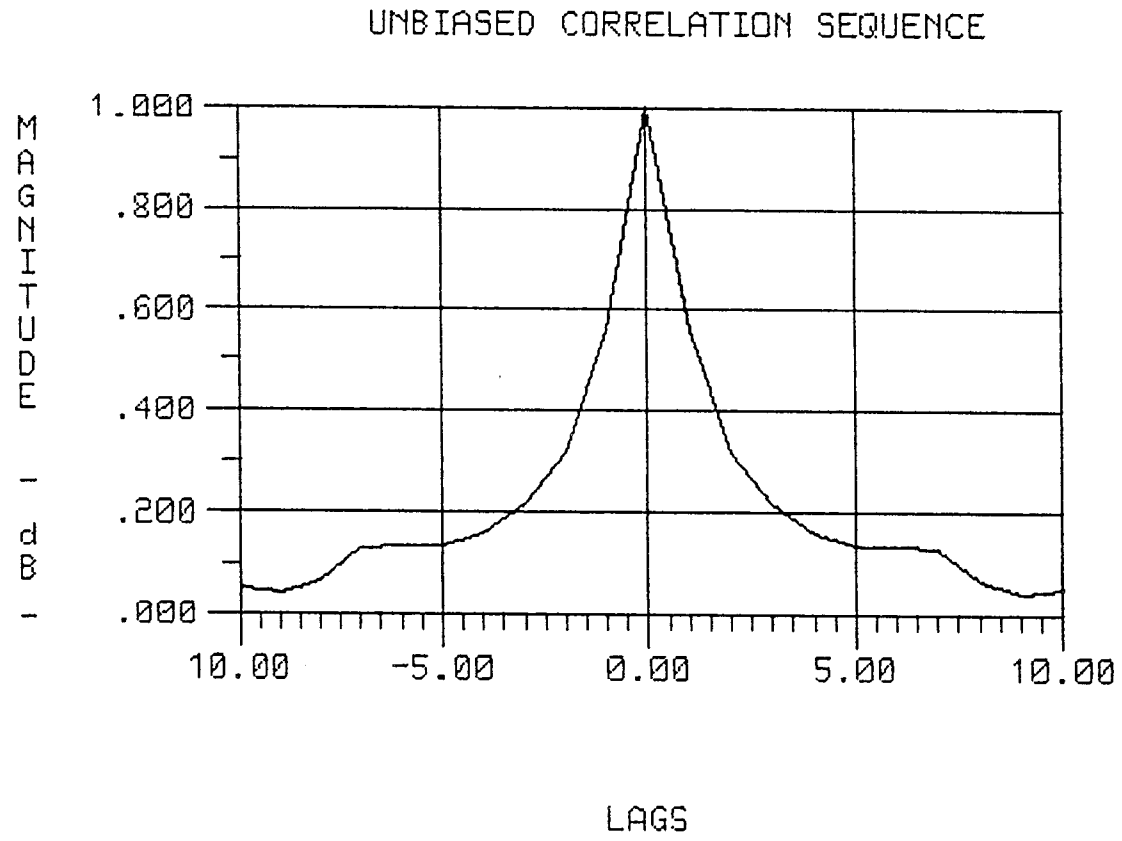


Figure 21. Correlation Sequence Estimate of Bin 11 Data.

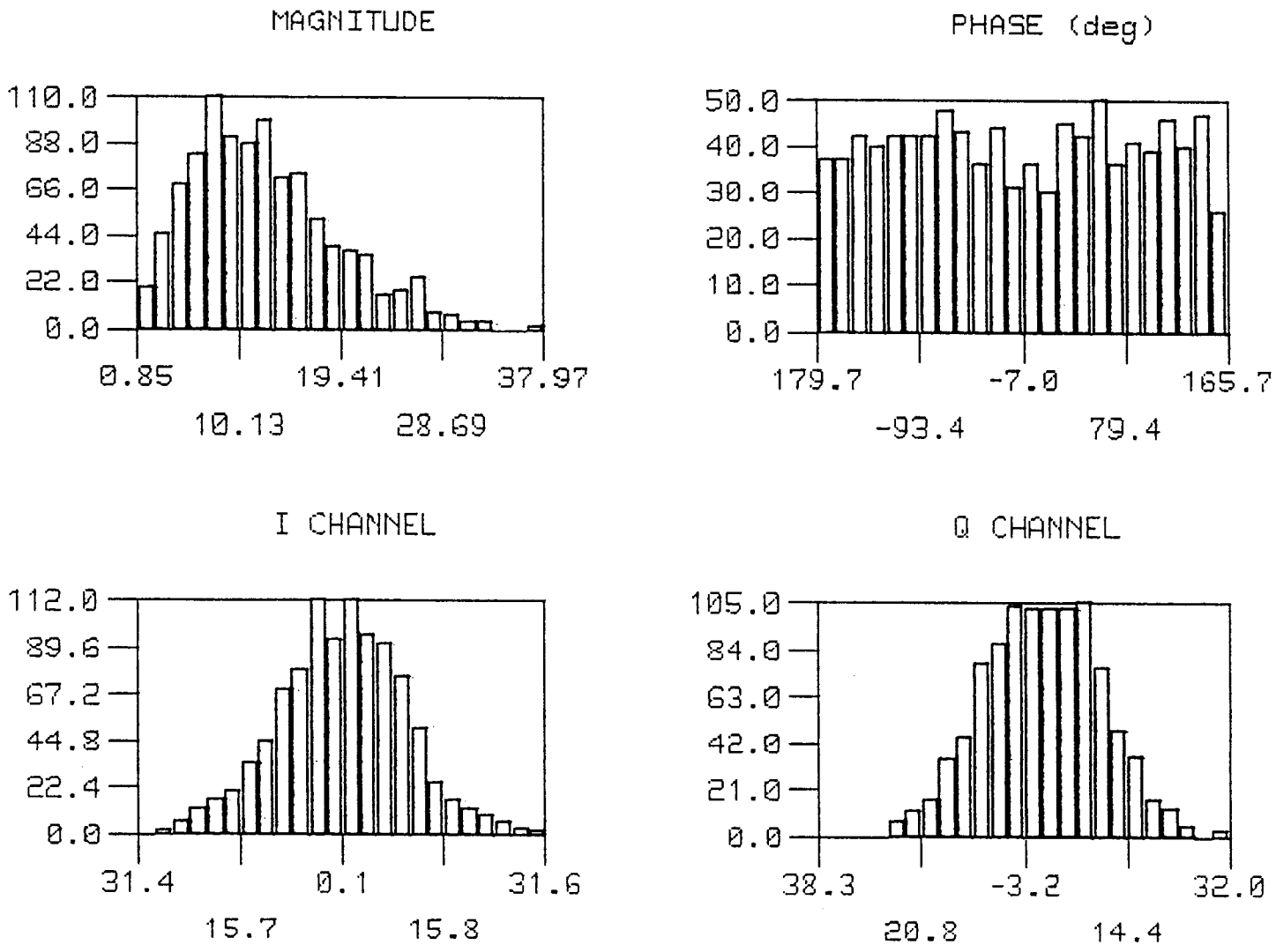


Figure 22. Histograms of Bin 11 Raw Data.

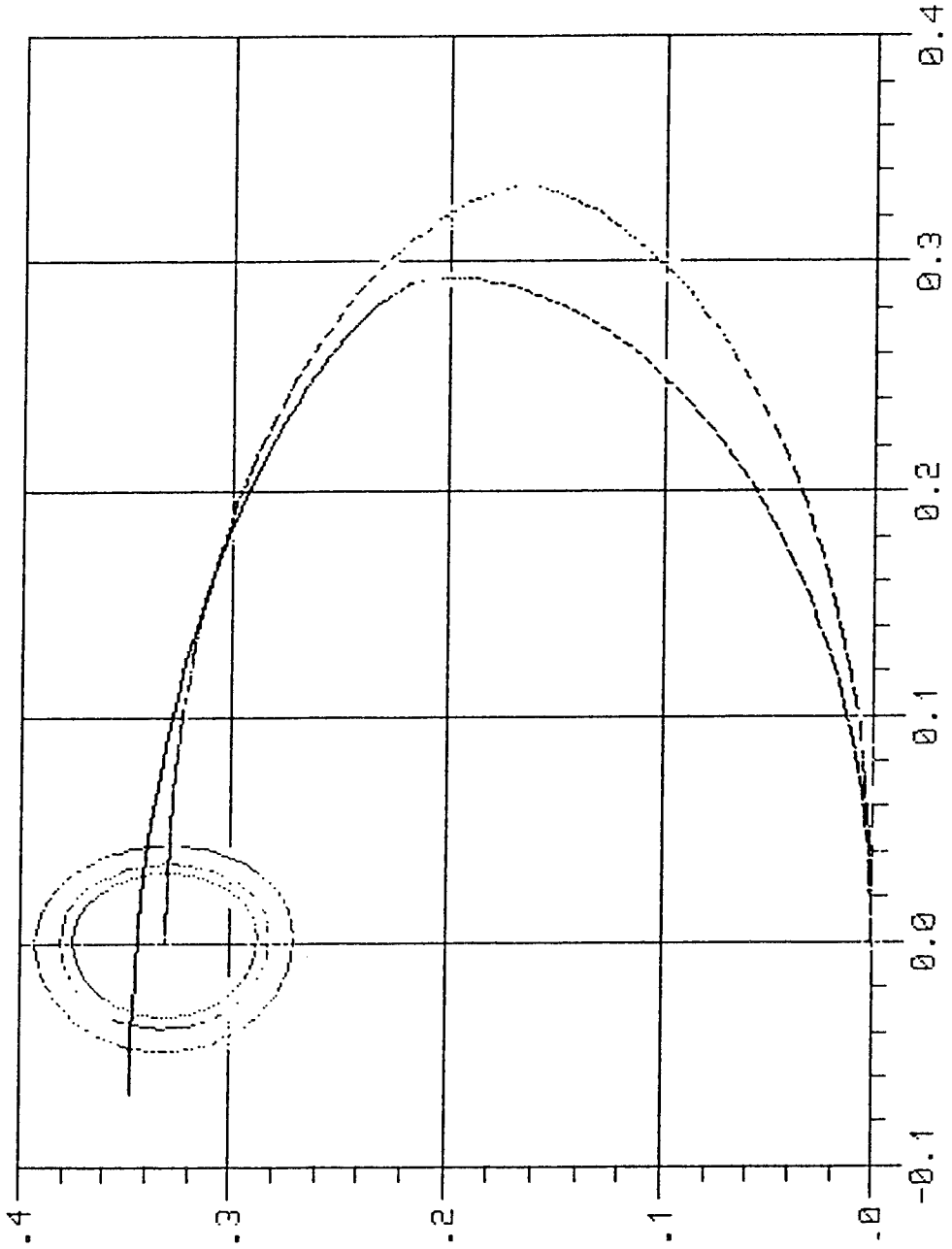
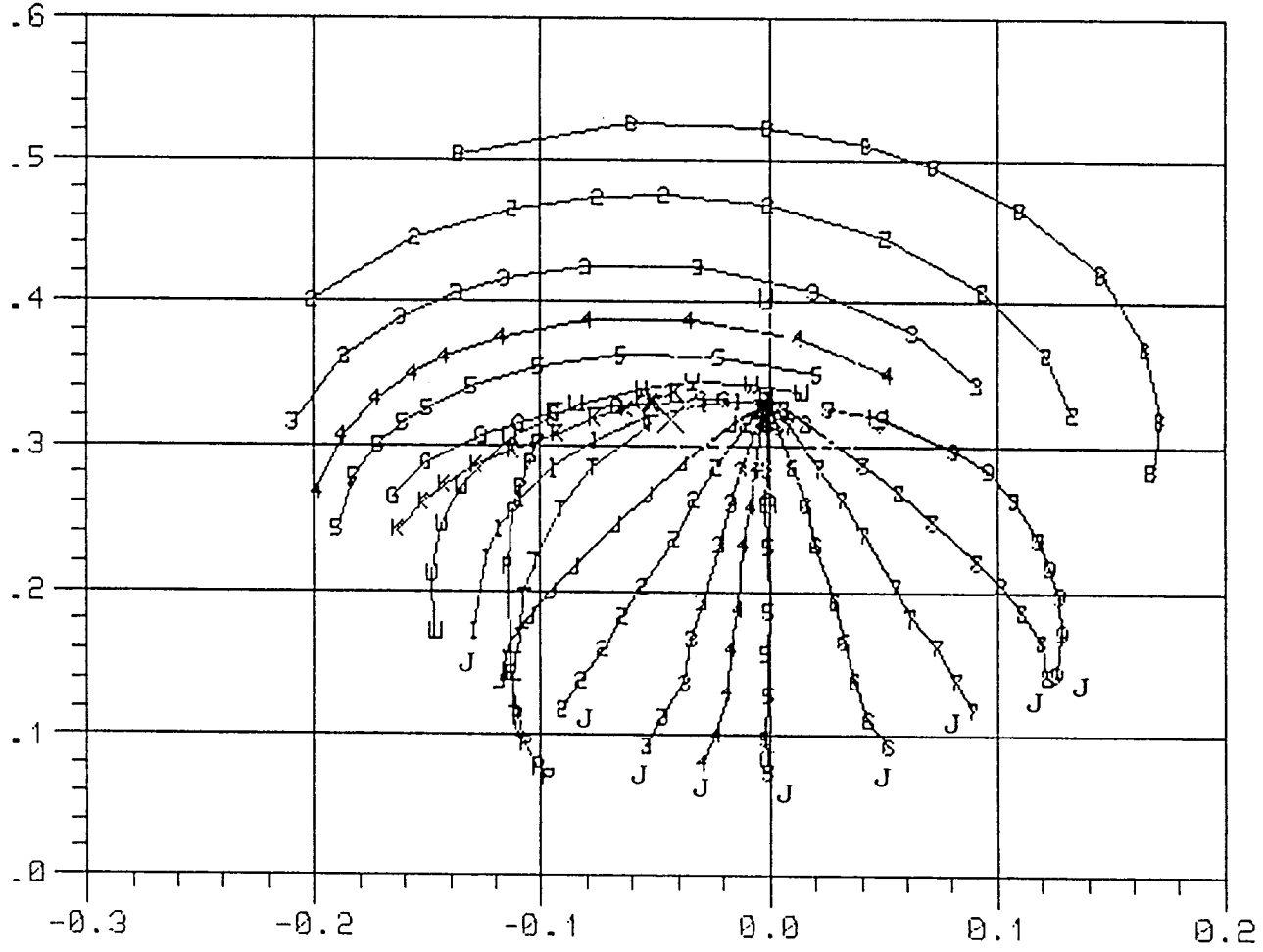


Figure 23. Goodness-Of-Fit Plot for Bin 11.



Nomenclature Indicates
Type of Distribution

B, 2, 3, 4, and 5
Beta Distribution
with different values
for the first shape
parameter

G – Gamma Distribution
K – K Distribution
W – Weibull Distribution
P – Pareto Distribution
T – Extreme Value
Distribution
L – LogNormal Distribution
U – Uniform Distribution
E – Exponential
Distribution
J – SU-Johnson System
(family of 9 curves)

Figure 24. PDF Approximation Chart for Bin 11.

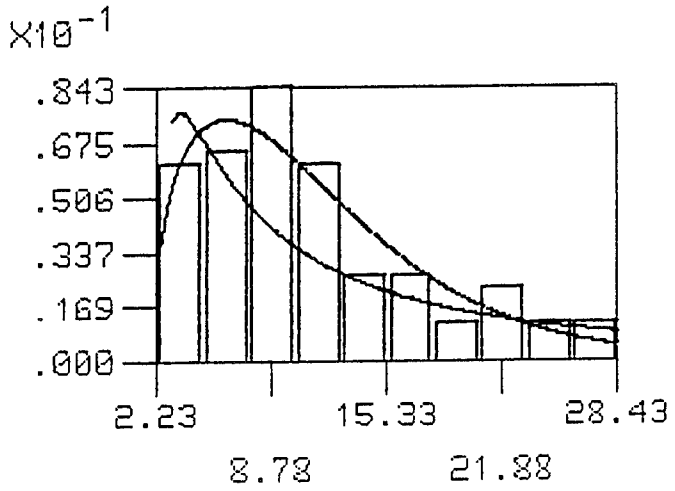


Figure 25. Overlay of Best/Worst PDF Approximations for Bin 11.

Clutter cell 12 results are presented in Figures 26 through 31. The time sequence of echo amplitudes in Figure 26 shows magnitudes closer to those of cells 8 -10 and higher than that of cell 11. The trend in phase is toward slower variations. The correlation sequence shown in Figure 27 has a structure similar to that of cell 11, possessing a tail of higher magnitude than previous cells. The decorrelation time of this cell is estimated to be 7.5 lags. Presentation of the data in histogram form in Figure 28 shows the general magnitude shape to be typical of the observed clutter cells, but the phase departs from its characteristic near-uniform behavior. This may be due to the clutter cell geometry defined by the high range resolution of the radar measurement system. With the small cell sizes, on the order of a few feet, the number and shape of scattering elements composing the cell may differ, thereby producing differences in temporal echoes. In the goodness-of-fit test of Figure 29, the termination of the sample data linked vector on the 0.05 confidence contour and the failure of its trajectory to closely follow that of the null hypothesis, leads to the conclusion that the sample data are not best described by the Gaussian distribution. This is verified by Figure 30, where the data termination is closest to the Weibull curve. Also, Figure 31 provides the visual comparison of the algorithm-chosen distribution with sample data.

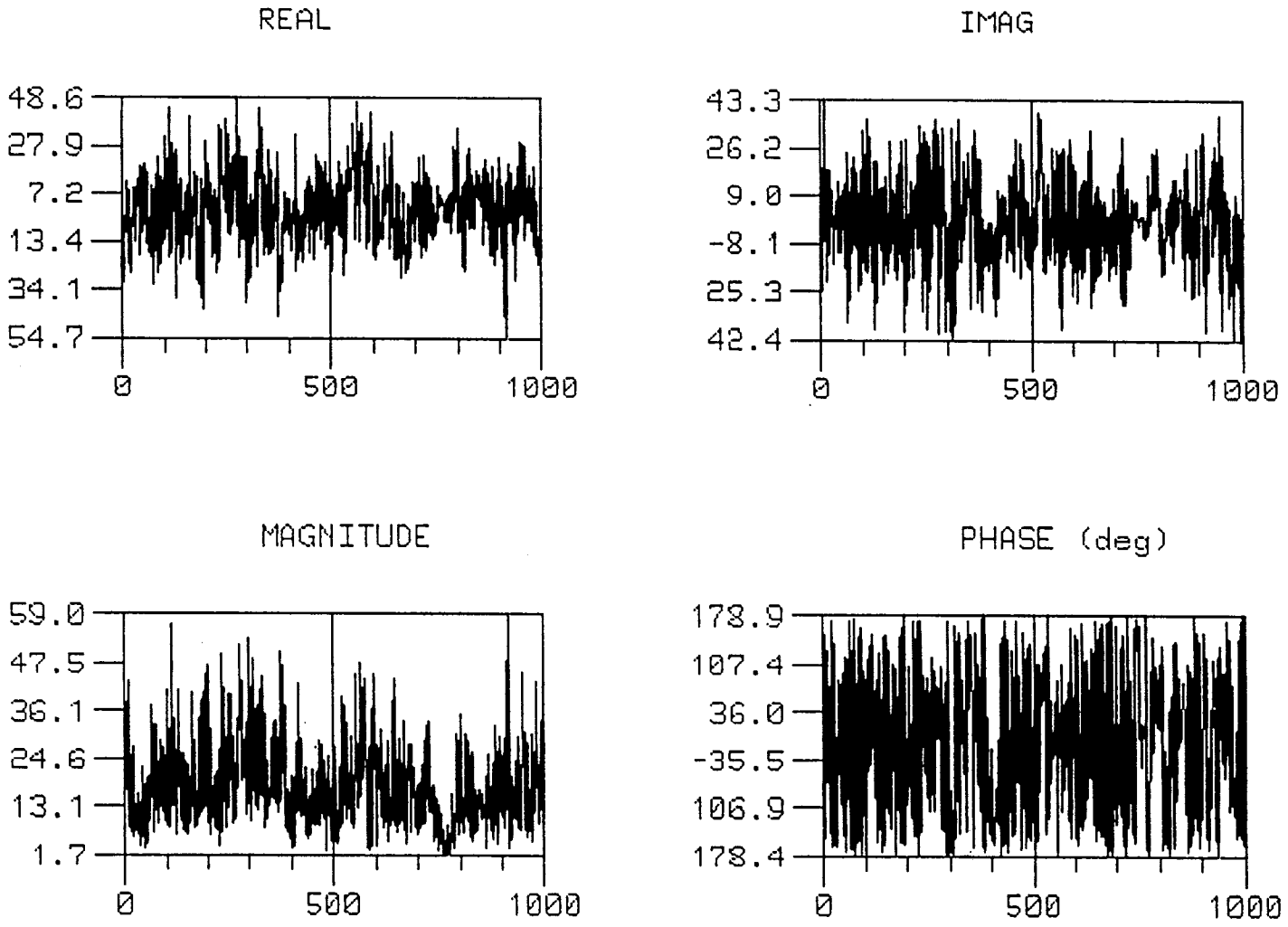


Figure 26. Raw Data of Bin 12.

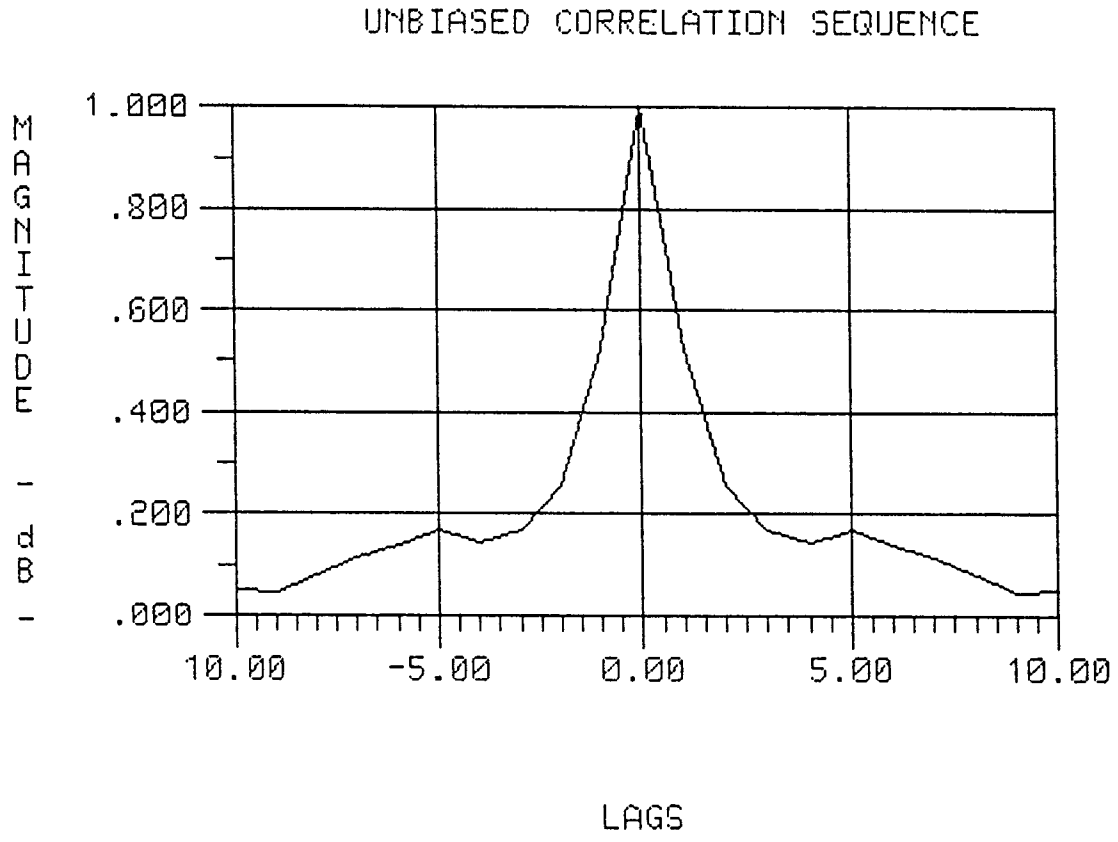


Figure 27. Correlation Sequence Estimate of Bin 12 Data.

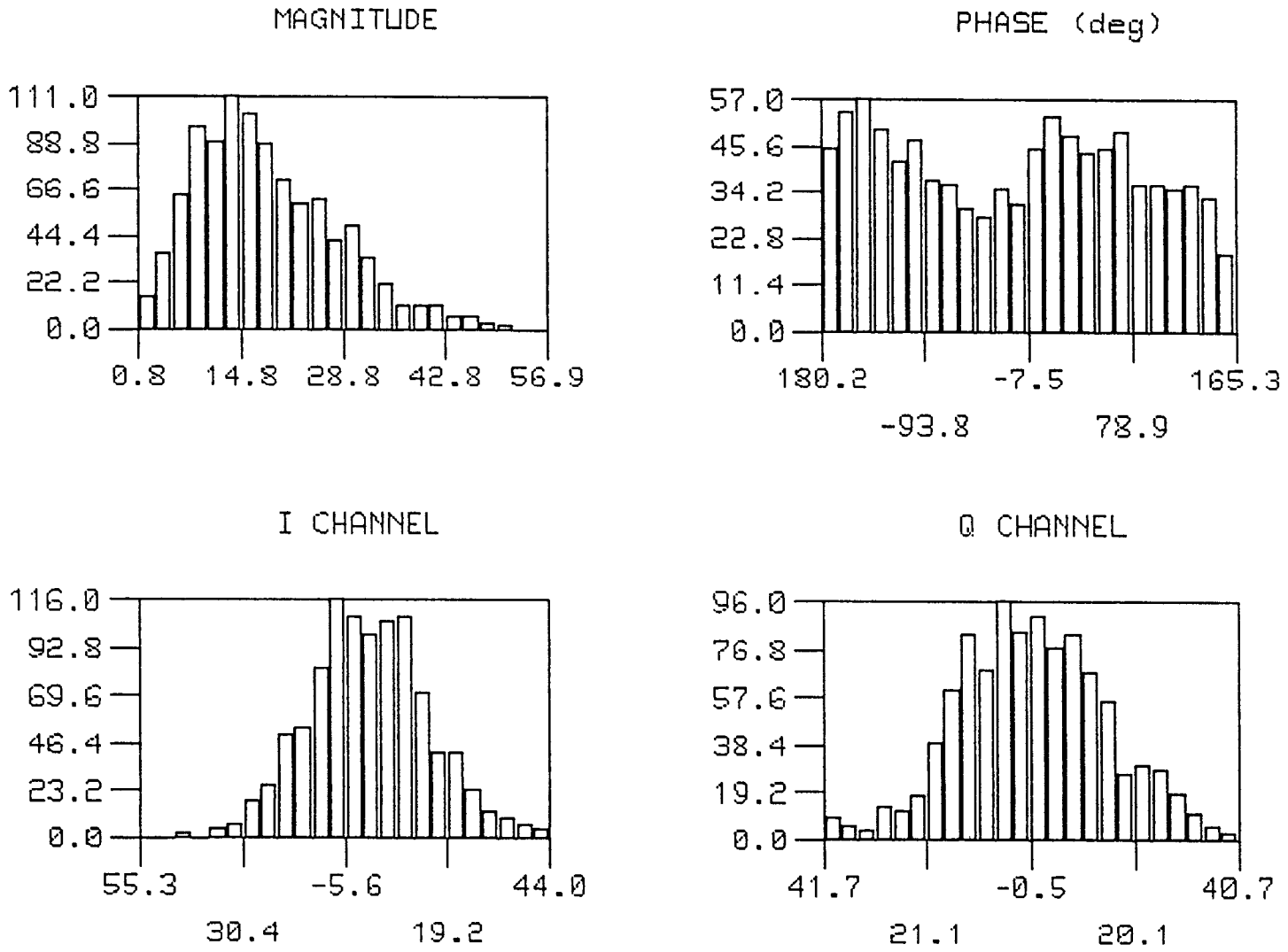


Figure 28. Histograms of Bin 12 Raw Data.

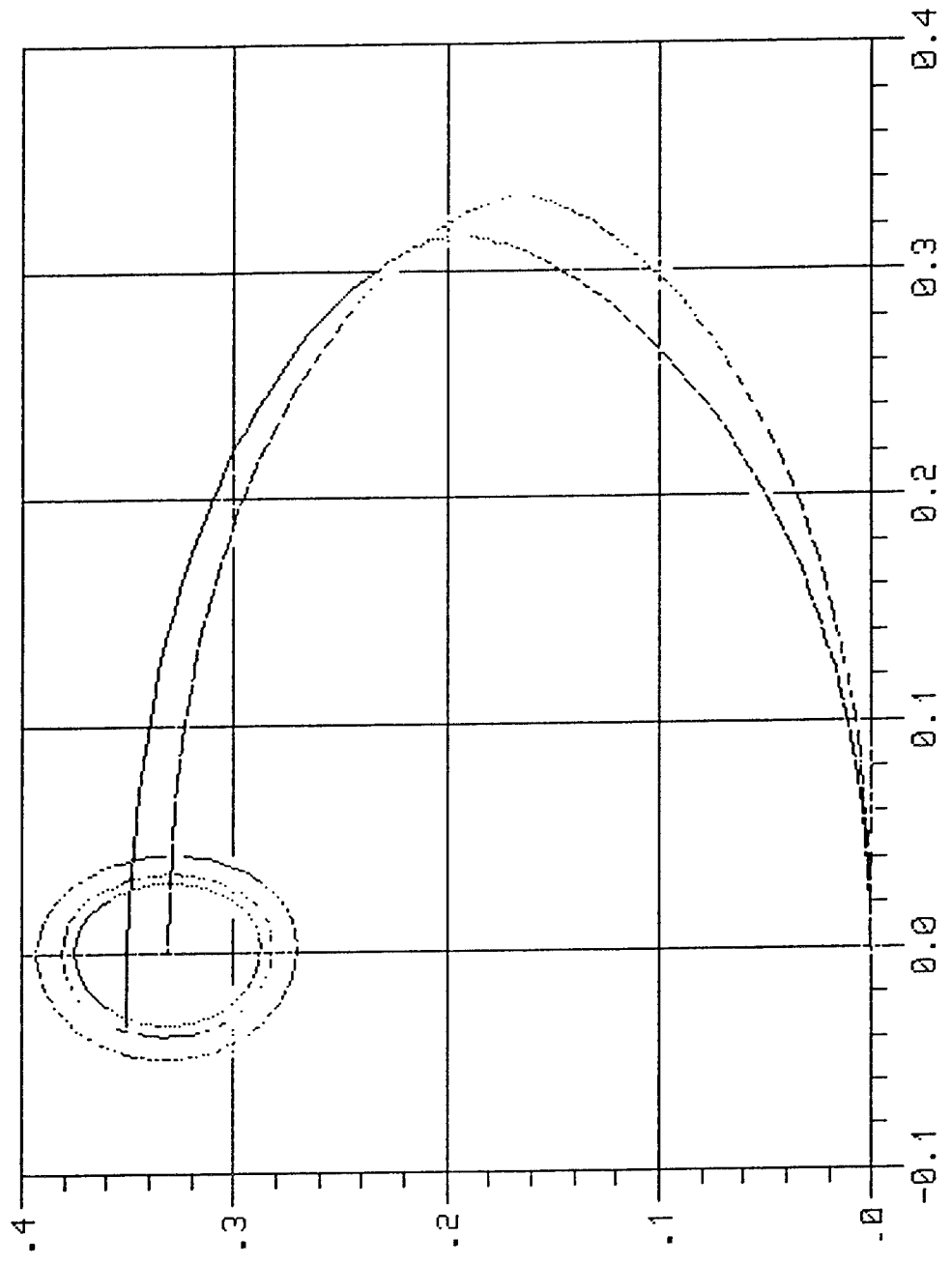
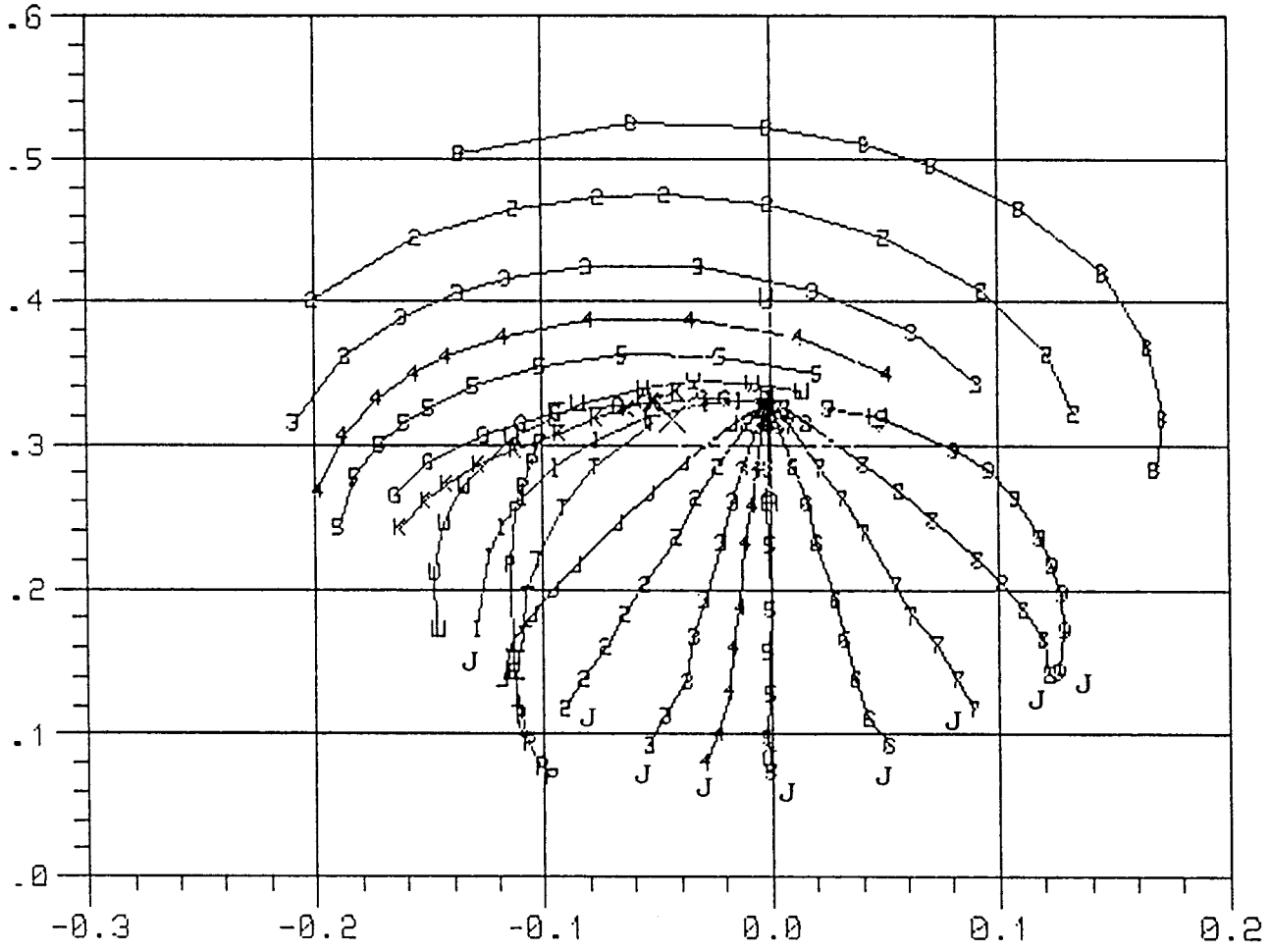


Figure 29. Goodness-Of-Fit Plot for Bin 12.



Nomenclature Indicates Type of Distribution

B, 2, 3, 4, and 5 – Beta Distribution with different values for the first shape parameter
G – Gamma Distribution
K – K Distribution
W – Weibull Distribution
P – Pareto Distribution
T – Extreme Value Distribution
L – LogNormal Distribution
U – Uniform Distribution
E – Exponential Distribution
J – SU-Johnson System (family of 9 curves)

Figure 30. PDF Approximation Chart for Bin 12.

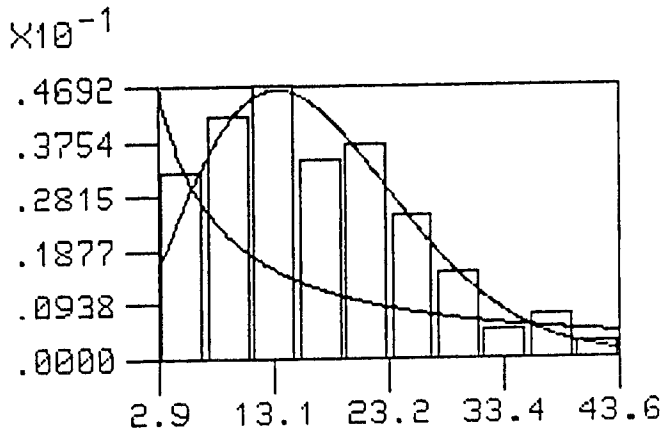


Figure 31. Overlay of Best/Worst PDF Approximations for Bin 12.

Figures 32 through 37 present analysis results of the clutter echoes from resolution cell 13. The raw data are shown in Figure 32, where the echo amplitudes are among the highest observed from this group of clutter cells. Several high-amplitude peaks are observed, instead of just a few occurrences. Also, the phase appears to be less rapidly varying, as was observed in some previous cells. The decorrelation time, determined from the correlation sequence estimate of Figure 33 is 8.5 lags, or 1.7 seconds; the longest time observed for these clutter cells. These differing correlation sequences and varying decorrelation times reinforce the hypothesis of different scattering mechanisms among neighboring resolution cells. The phase histogram of Figure 34 again appears to depart from the uniform case. Observation of the sample data trajectory and end-point in Figure 35 clearly dismisses the null hypothesis as a good descriptor of the data set. For this clutter cell, the amplitude statistics are determined to be best represented by a beta distribution, as shown in Figure 36. The overlay of closest and farthest candidate PDFs onto the data are shown in Figure 37.

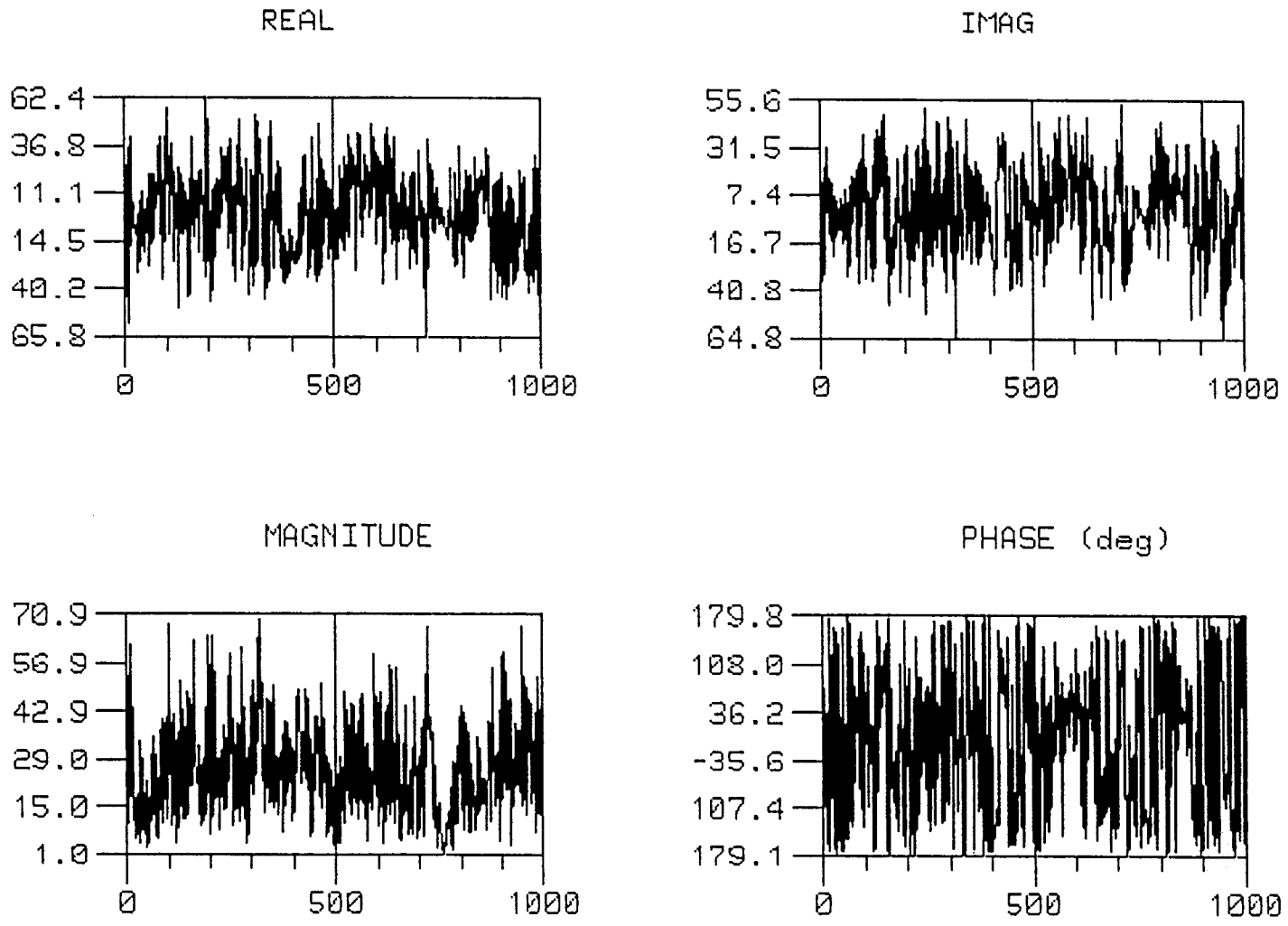


Figure 32. Raw Data of Bin 13.

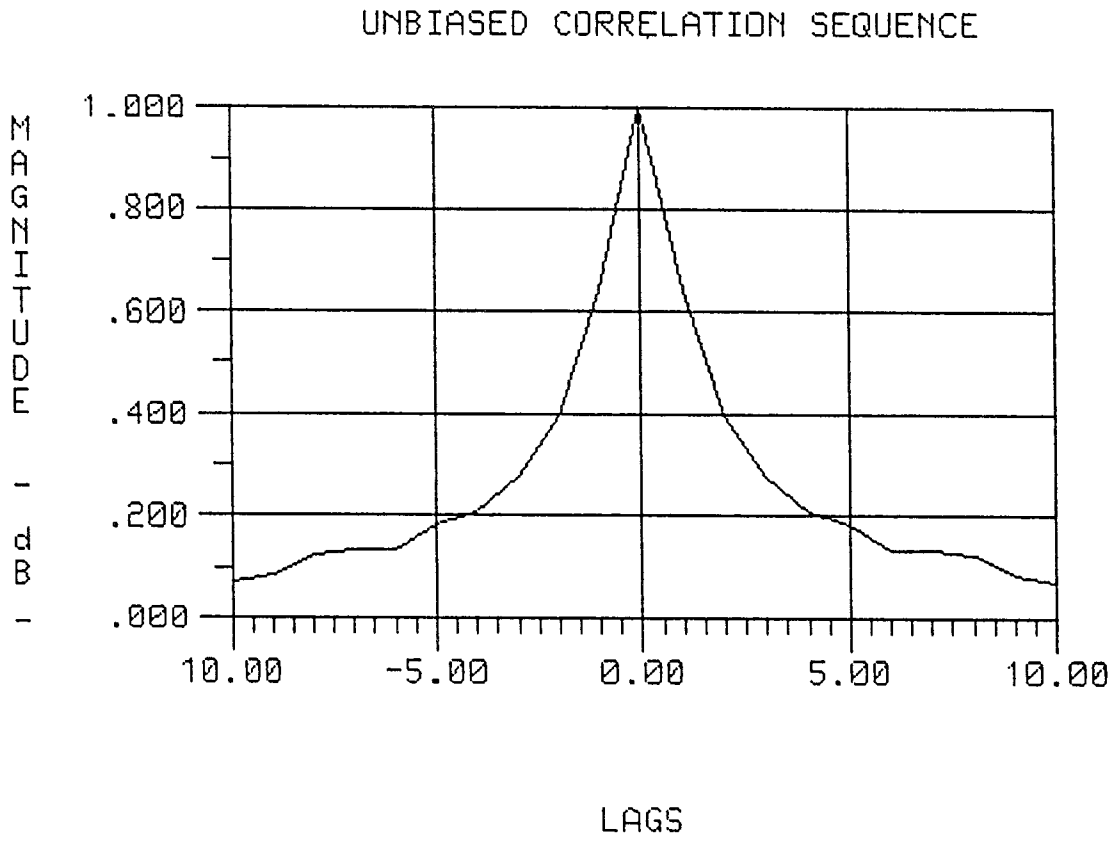


Figure 33. Correlation Sequence Estimate of Bin 13 Data.

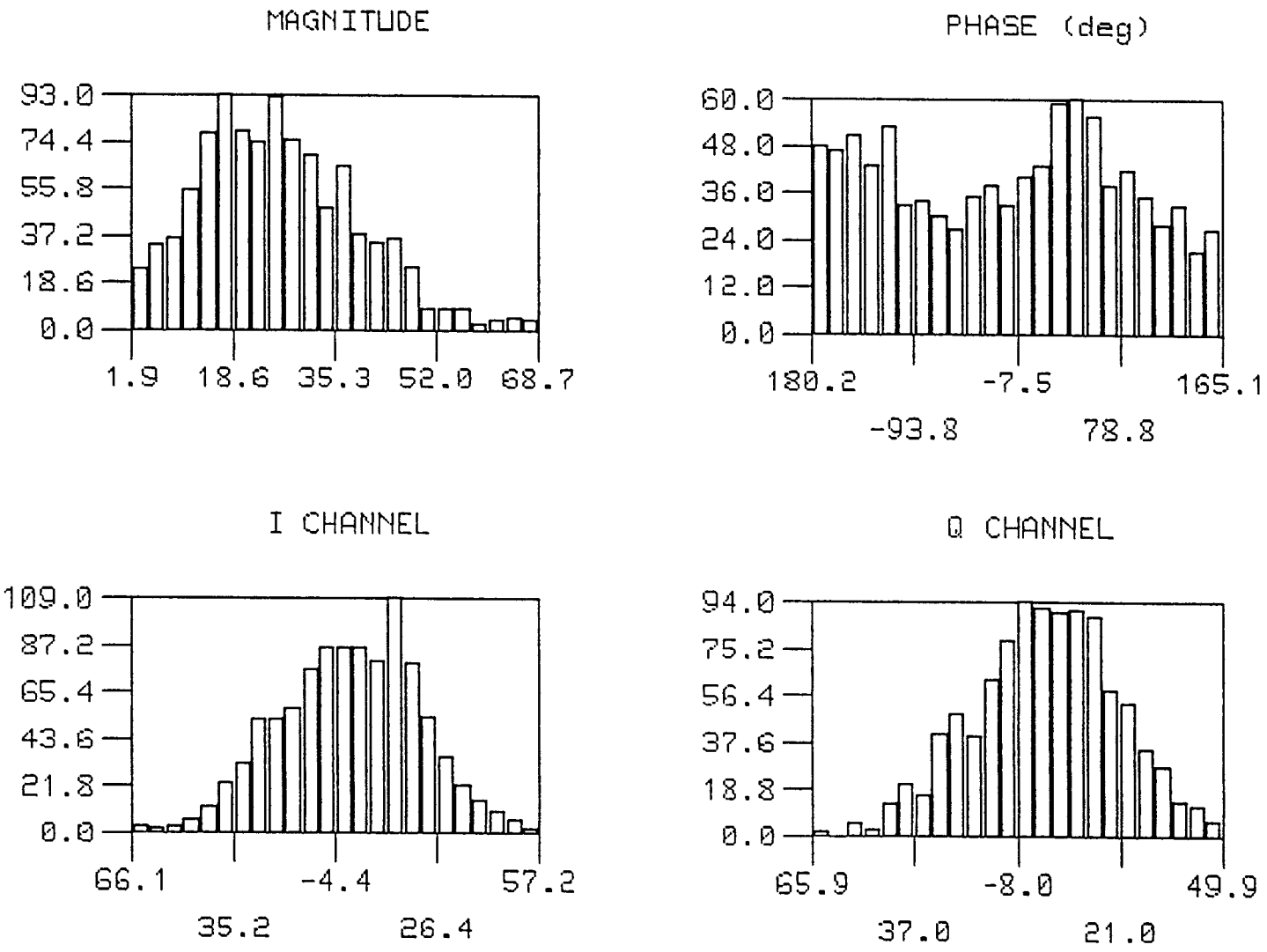


Figure 34. Histograms of Bin 13 Raw Data.

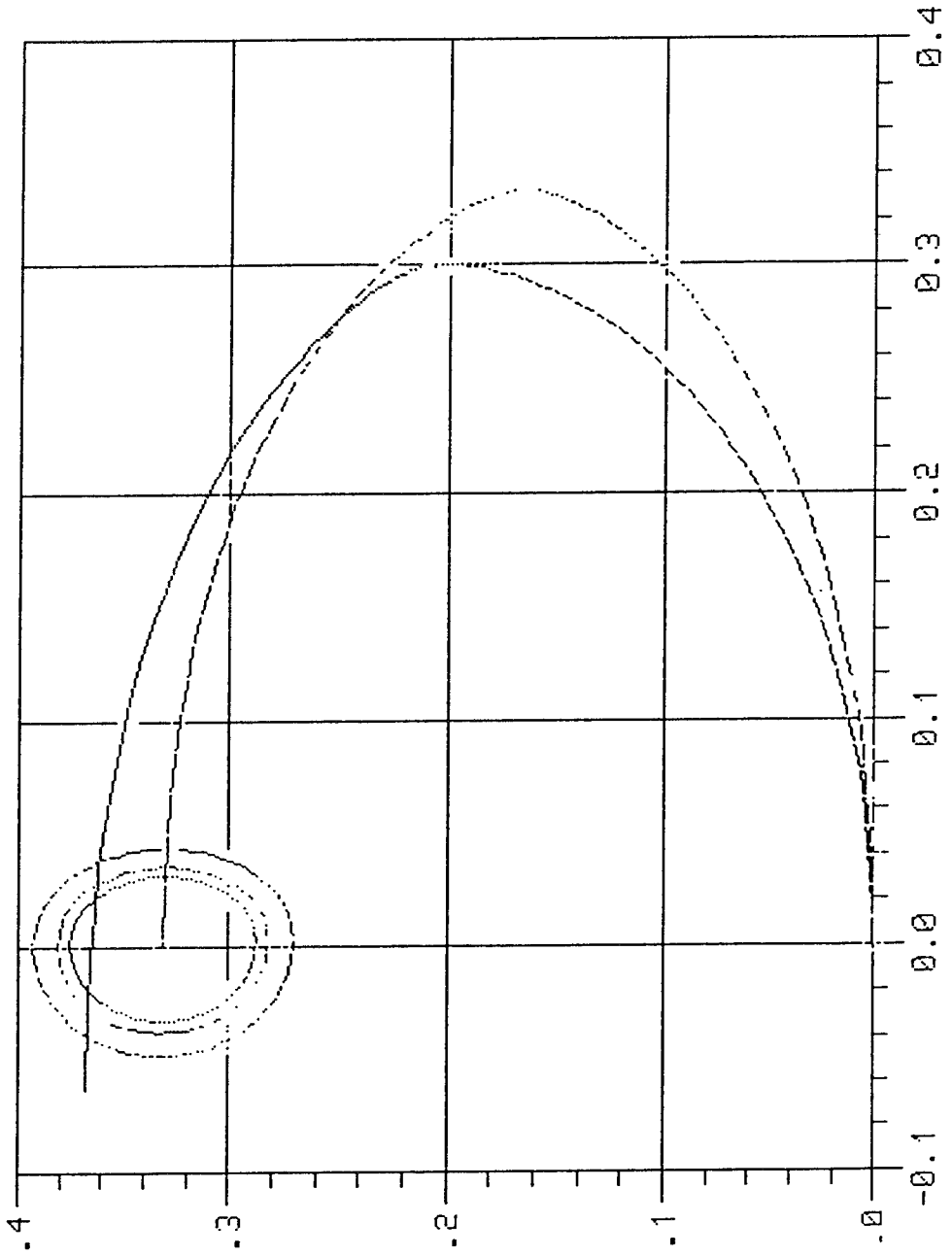
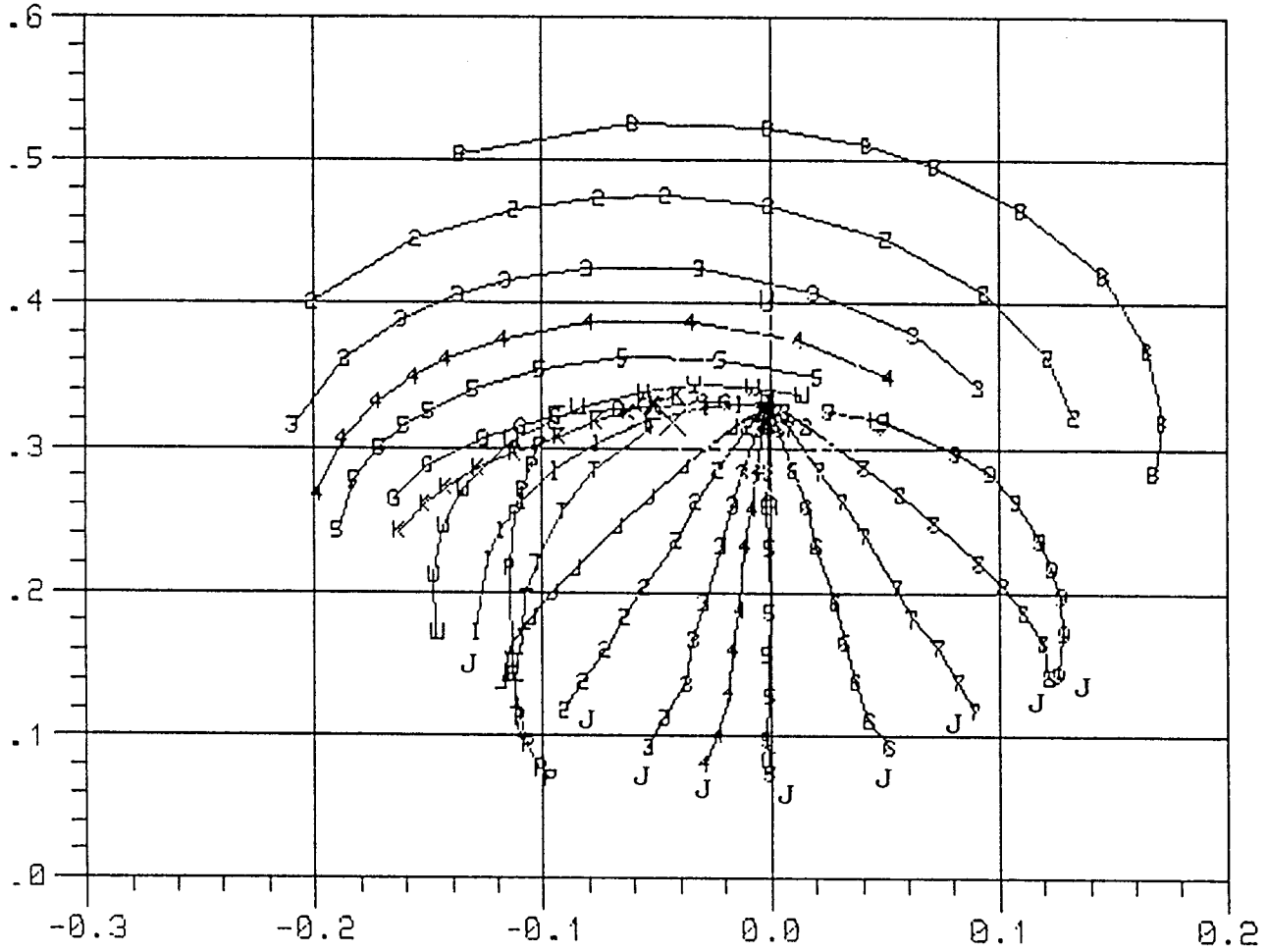


Figure 35. Goodness-Of-Fit Plot for Bin 13.



Nomenclature Indicates
Type of Distribution

B, 2, 3, 4, and 5
Beta Distribution with different values for the first shape parameter

G – Gamma Distribution
K – K Distribution
W – Weibull Distribution
P – Pareto Distribution
T – Extreme Value Distribution
L – LogNormal Distribution
U – Uniform Distribution
E – Exponential Distribution
J – SU-Johnson System (family of 9 curves)

Figure 36. PDF Approximation Chart for Bin 13.

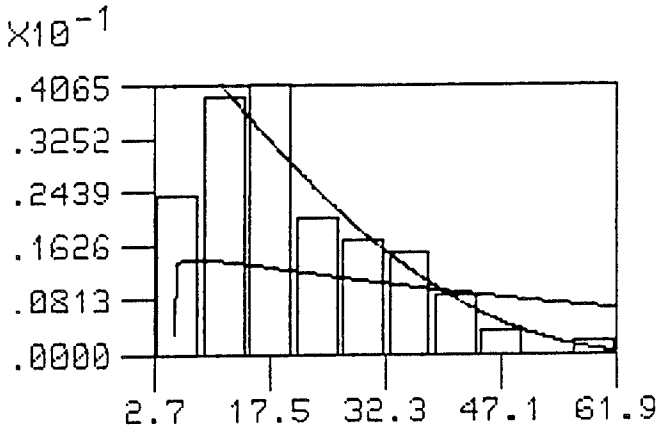


Figure 37. Overlay of Best/Worst PDF Approximations for Bin 13.

In Figure 38 the time history of amplitude echoes of cell 14 shows a single peak close to that of the previous cell, but the general trend is toward lower amplitudes. The phase behavior is similar with both fast and slower variations. The correlation sequence of Figure 39 shows that this cell has the greatest decorrelation time - greater than 10 lags. Histograms of both amplitude and phase in Figure 40 show trends similar to the previous cell. The sample data are determined not to be characterized best by the null hypothesis PDF, since similar trajectories and close termination points are required to satisfy the goodness-of-fit test. As seen in Figure 41 these conditions are not met. Instead, data from this clutter cell are best described by the beta distribution of the previous cell, but with a different shape parameter. This is illustrated in Figure 42. Comparison of the estimated PDF with experiment data is shown in Figure 43. Note that the approximating PDF for cells 13 and 14 arise from the same trajectory corresponding to the beta distribution. However, observe the minor difference in the values of one of the shape parameters.

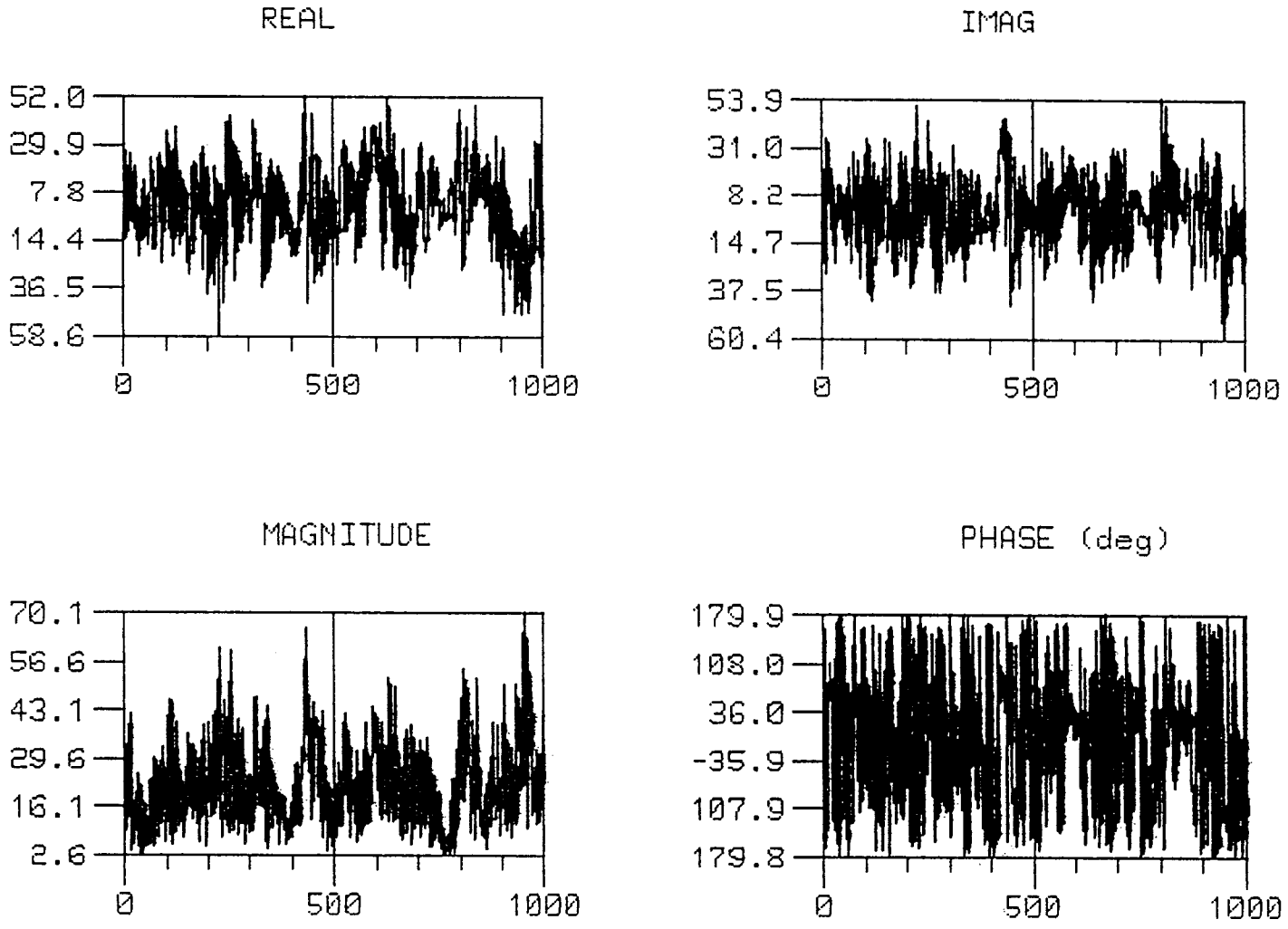


Figure 38. Raw Data of Bin 14.

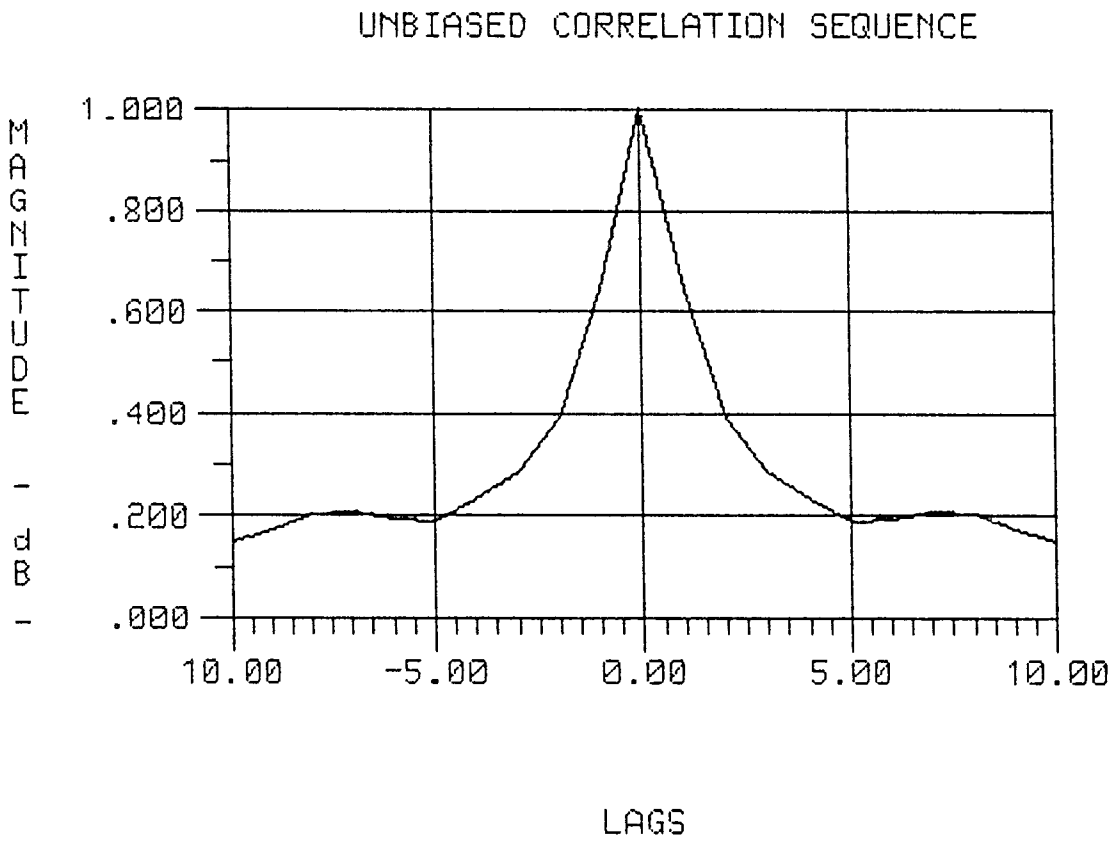


Figure 39. Correlation Sequence Estimate of Bin 14 Data.

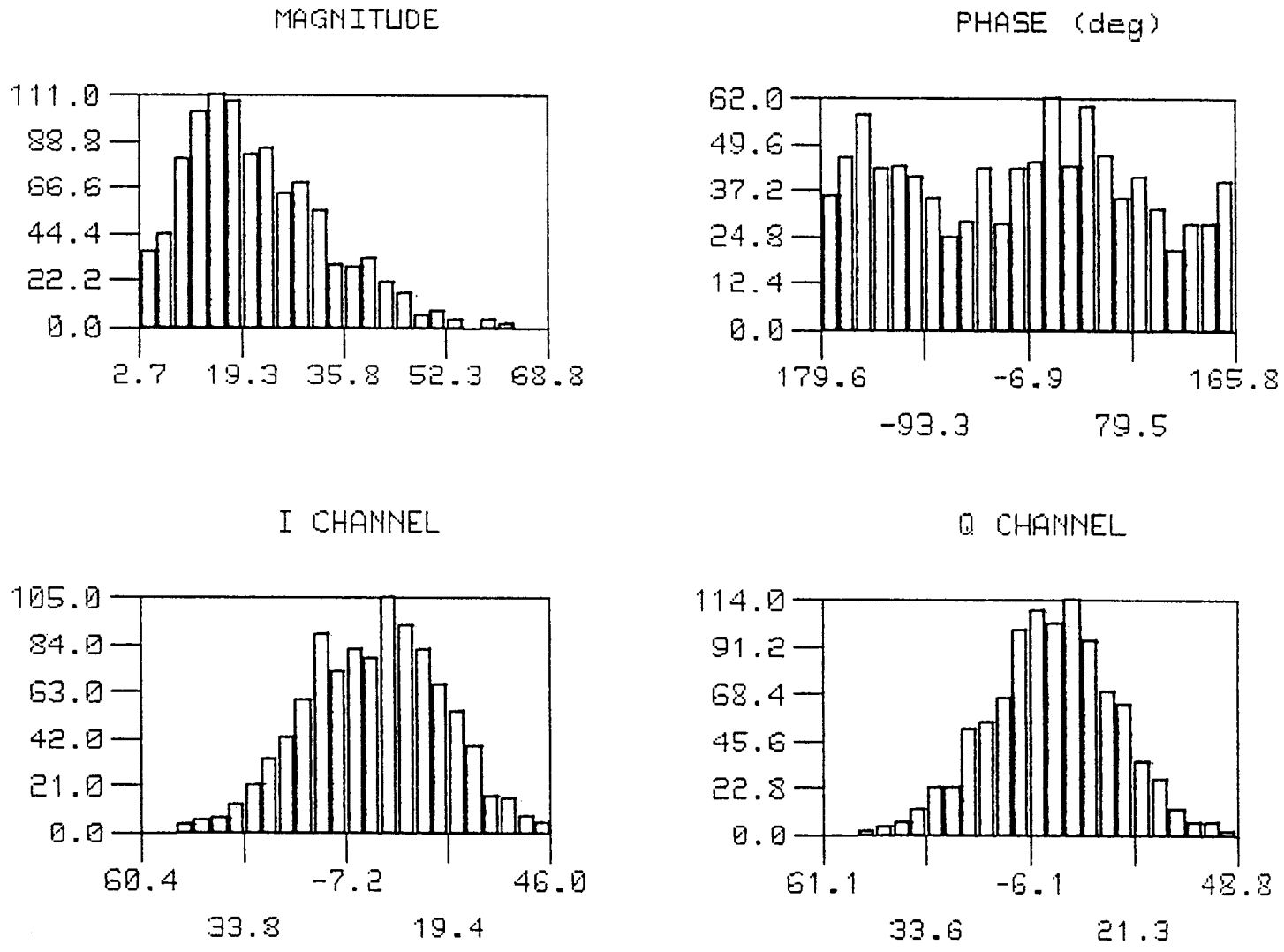


Figure 40. Histograms of Bin 14 Raw Data.

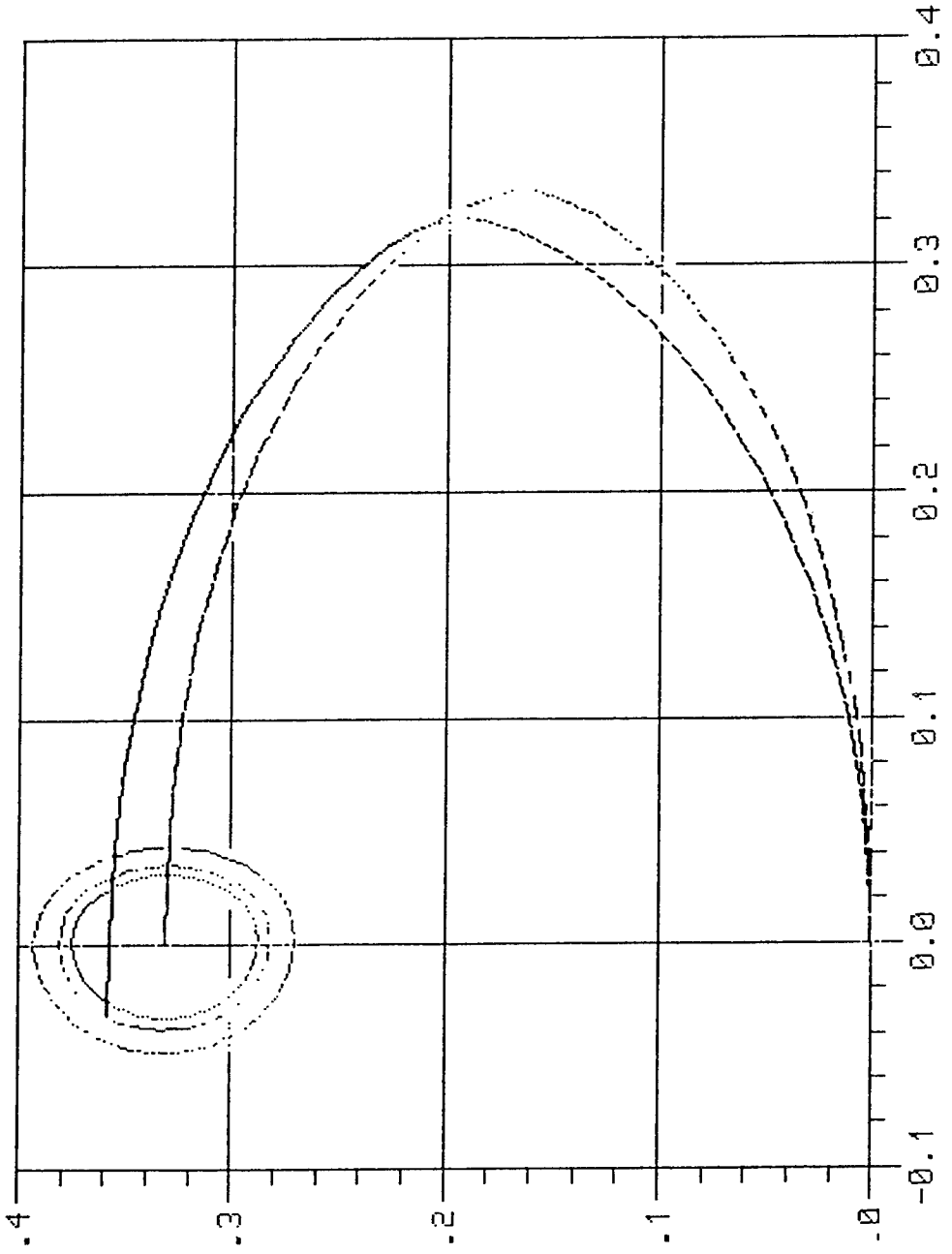
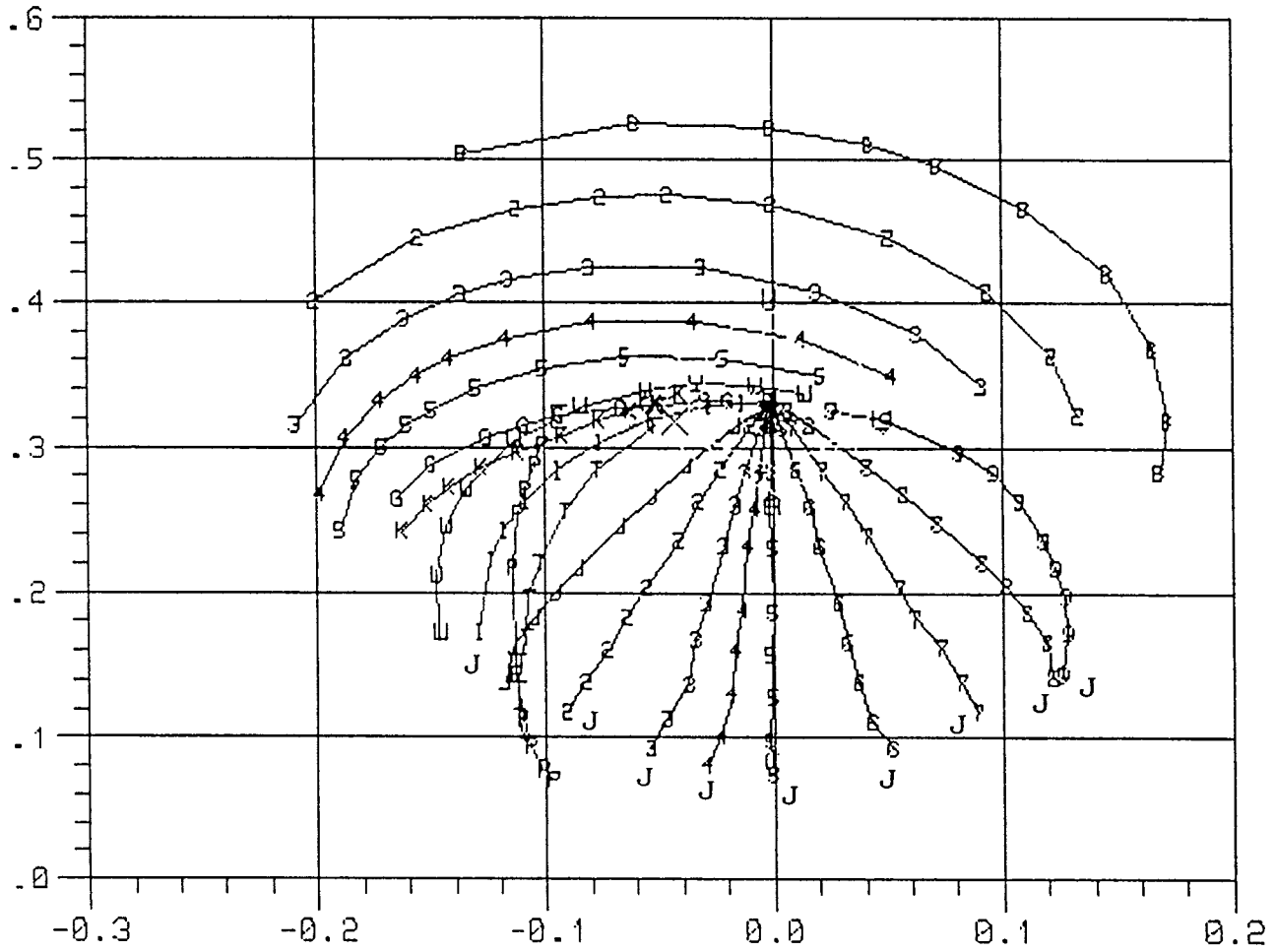


Figure 41. Goodness-Of-Fit Plot for Bin 14.



Nomenclature Indicates Type of Distribution

B, 2, 3, 4, and 5 – Beta Distribution with different values for the first shape parameter
G – Gamma Distribution
K – K Distribution
W – Weibull Distribution
P – Pareto Distribution
T – Extreme Value Distribution
L – LogNormal Distribution
U – Uniform Distribution
E – Exponential Distribution
J – SU-Johnson System (family of 9 curves)

Figure 42. PDF Approximation Chart for Bin 14.

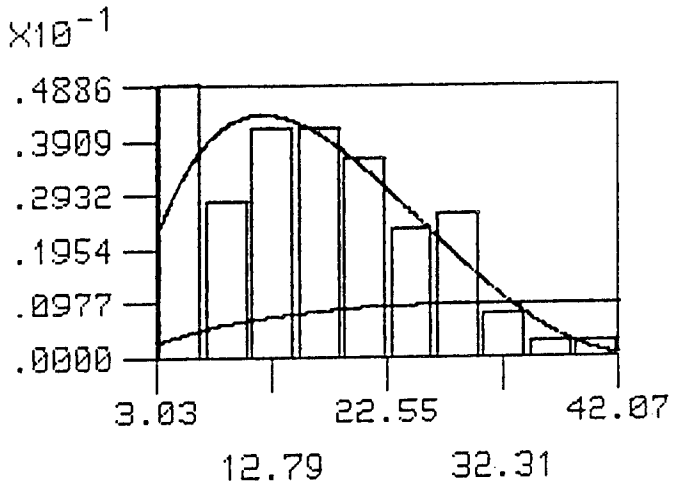


Figure 43. Overlay of Best/Worst PDF Approximations for Bin 14.

Resolution cell 15 possessed the highest magnitude clutter echoes of any of the observed cells, as illustrated in Figure 44. The decorrelation time as seen from the correlation estimate of Figure 45 was greater than 10 lags. The data in the histogram of Figure 46 had structures similar to cells 13 and 14, but as Figures 47 and 48 show, the data are better described by the Weibull distribution as in cells 9 through 11. The final visual assessment, comparing best and worst estimates of the data is shown in Figure 49.

The final cell in the set was number 16, which had the amplitude and phase fluctuations shown in Figure 50. The echo amplitudes of this cell decreased to the level of cells 9 through 12. However, the correlation estimate of Figure 51 shows that the decorrelation time is still greater than 10 lags, as with the previous two cells. The trends illustrated in Figure 52 are not greatly dissimilar to those of the closest neighbor cells, with the exception of the broader hump in the magnitude distribution. A look at the goodness-of-fit test in Figure 53 leads to the immediate rejection of the Gaussian distribution as best describing the amplitude statistics. Instead, the Weibull is again chosen as the better fit as seen in the PDF approximation chart of Figure 54. This is also verified by the best/worst overlay onto the data histogram in Figure 55.

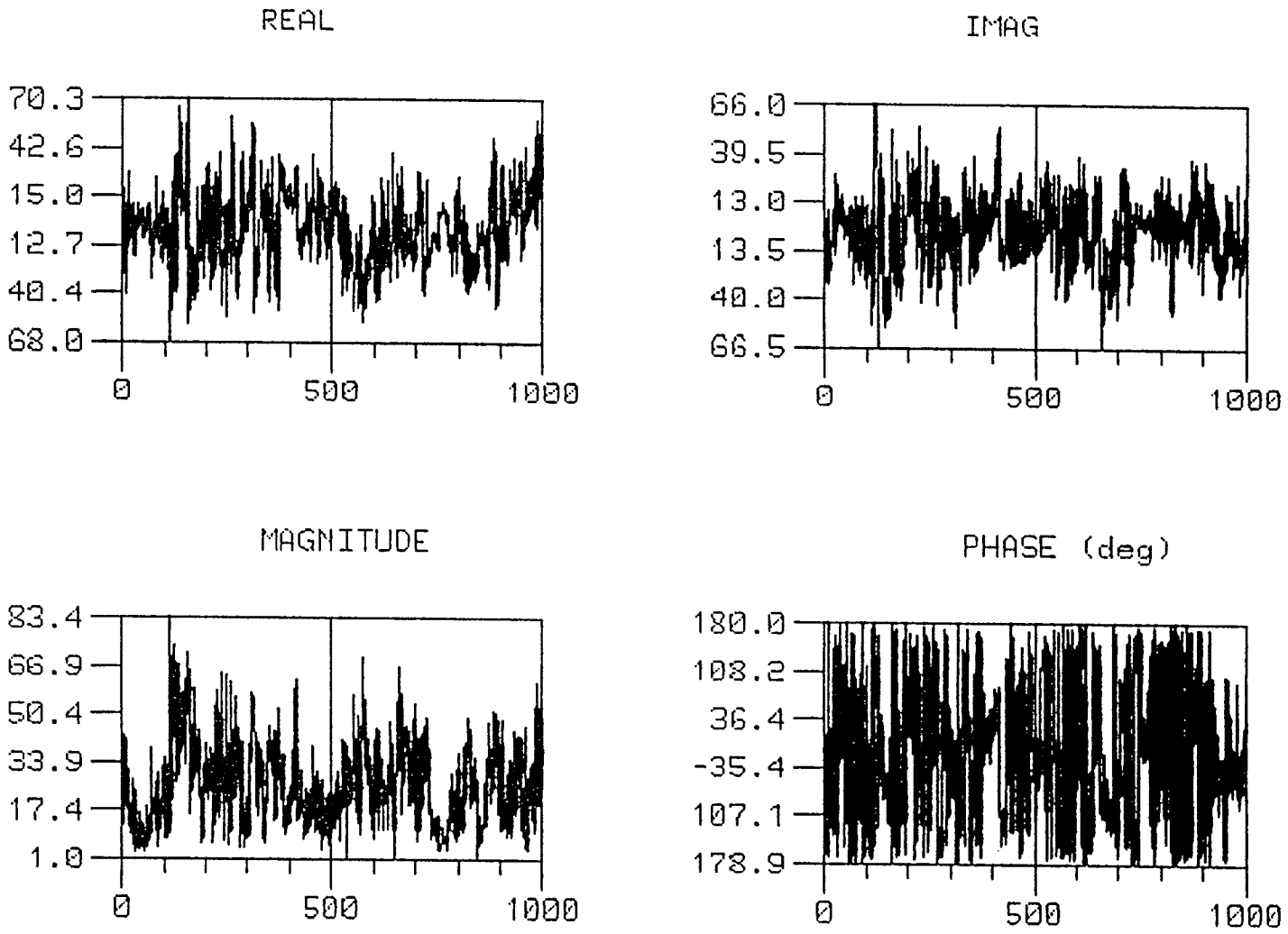


Figure 44. Raw Data of Bin 15.

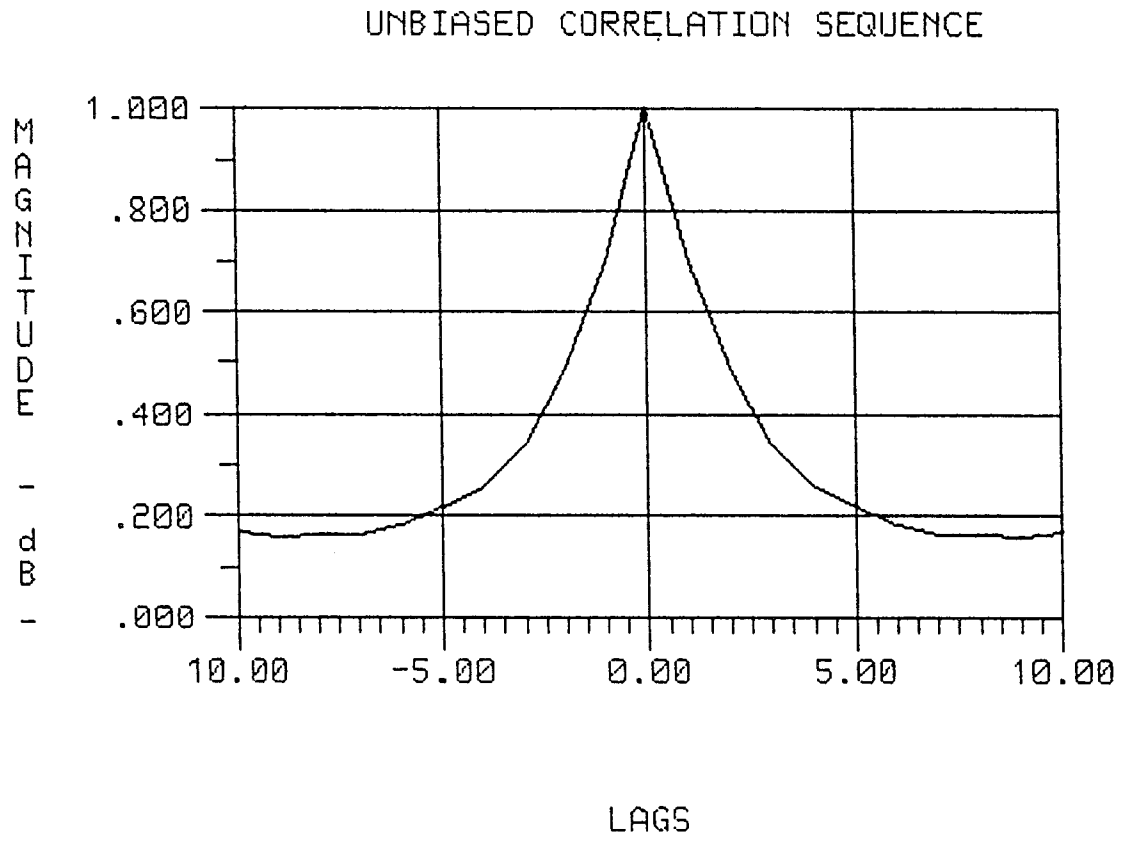


Figure 45. Correlation Sequence Estimate of Bin 15 Data.

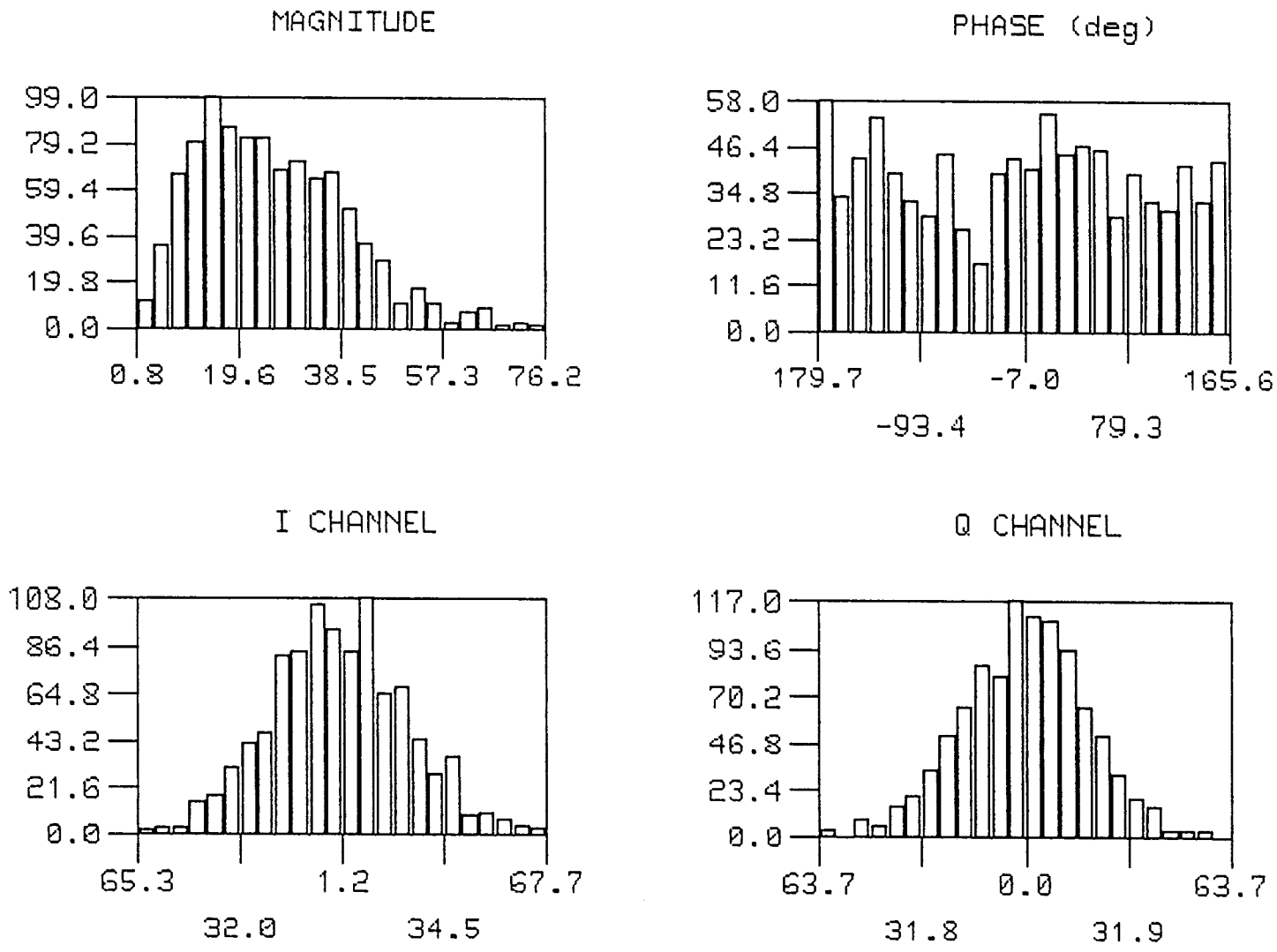


Figure 46. Histograms of Bin 15 Raw Data.

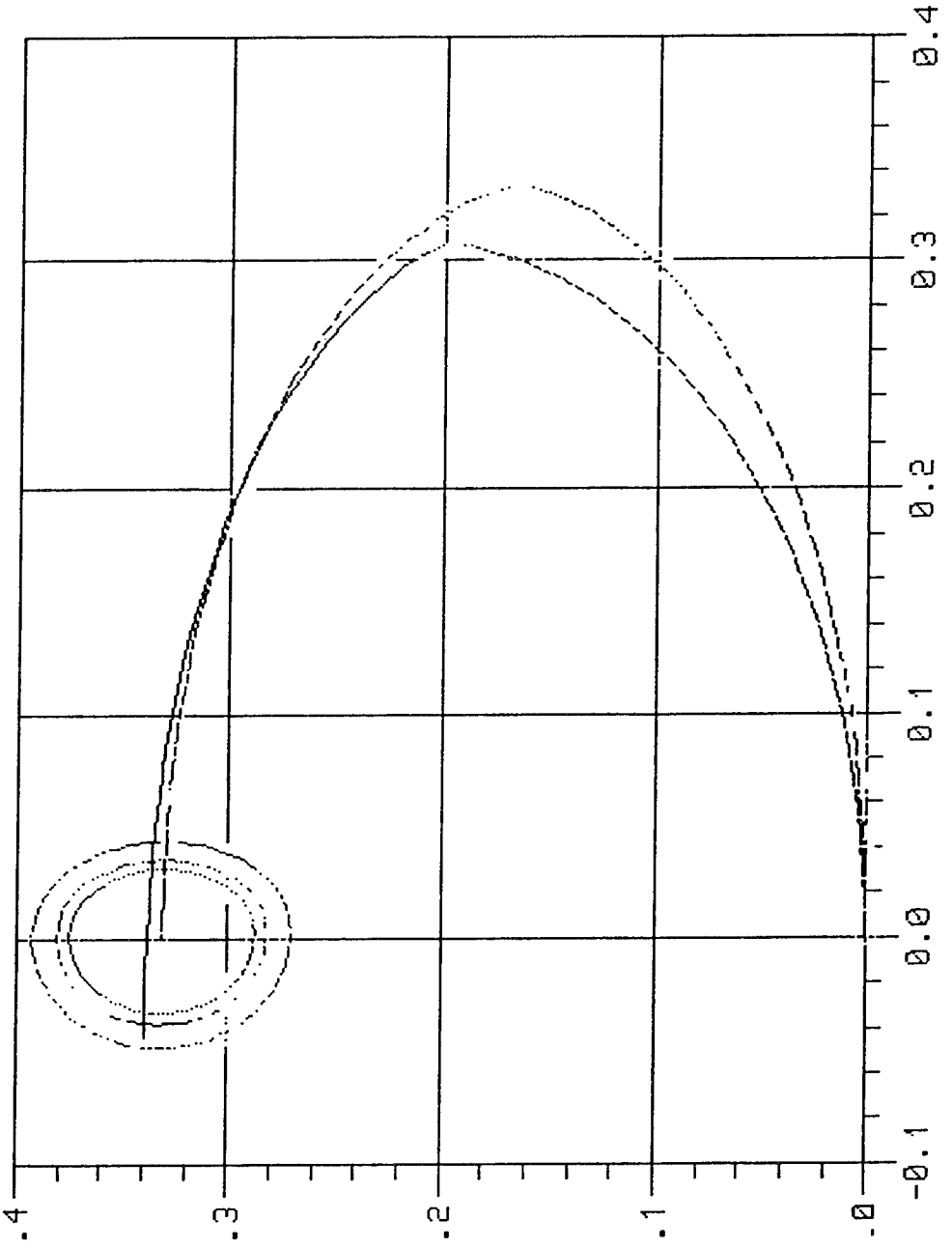


Figure 47. Goodness-Of-Fit Plot for Bin 15.

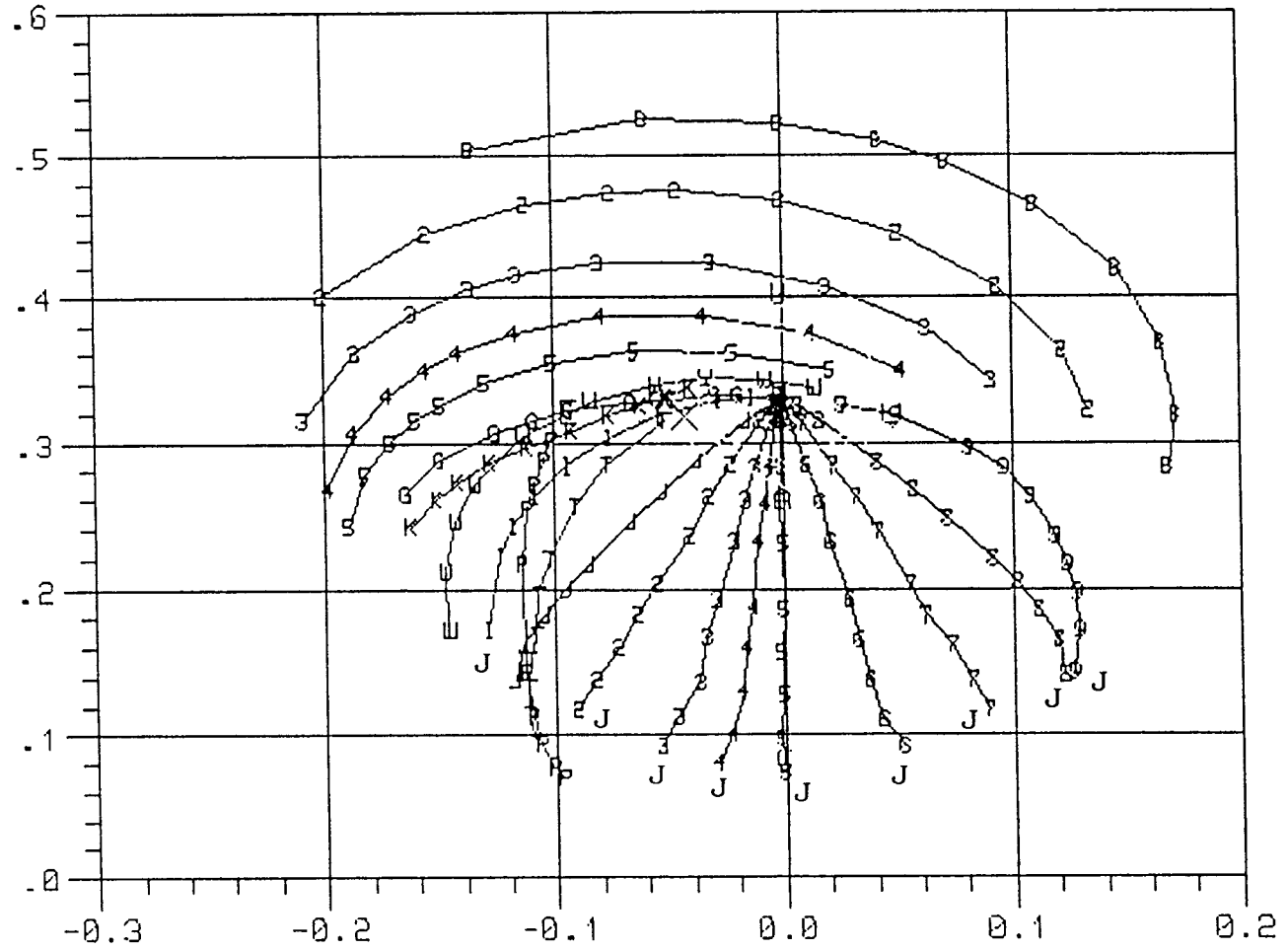


Figure 48. PDF Approximation Chart for Bin 15.

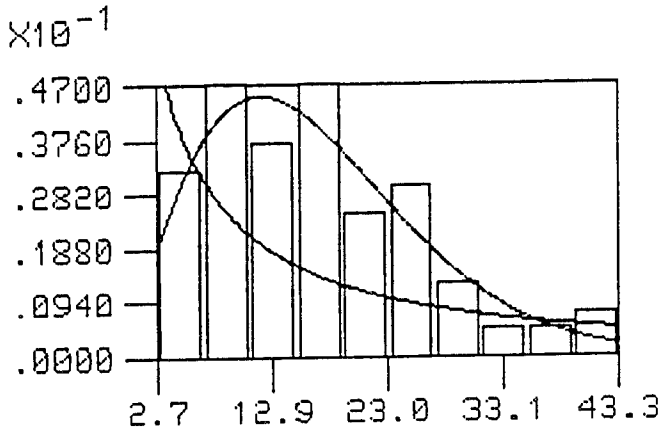


Figure 49. Overlay of Best/Worst PDF Approximations for Bin 15.

5. SUMMARY

Observations of the polarimetric properties of these same sets of data revealed a clustering of polarization state among neighboring range bins. In particular, bins 9-11,12-14, and 15-16 had similar polarization states. This result is corroborated by the fact that the best approximating amplitude probability density function chosen by the algorithm was Weibull for bins 9-11, beta for bins 12-14, and Weibull for bins 15-16. The shape parameters of the Weibull approximations exhibited minor variations for bins 9-11 and 15-16. Similar trends were observed for the shape parameter estimates of the beta distribution.

Future research should include the use of this algorithm for determining statistical properties of radar clutter data of different terrain types, resolution cell sizes and with different polarizations. Further, the algorithm should be used to determine the statistical properties pertaining to the spatial variation of bistatic radar clutter.

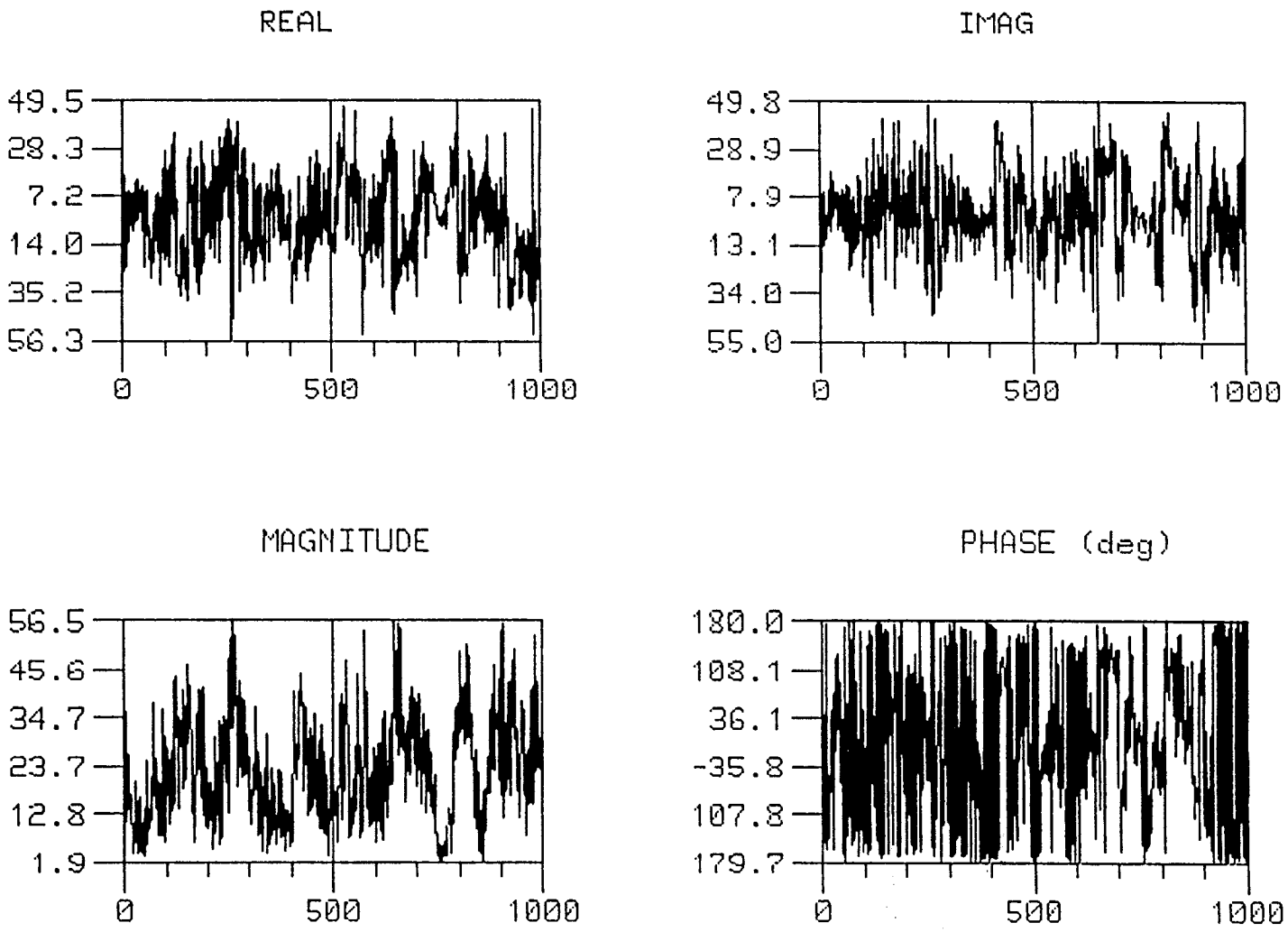


Figure 50. Raw Data of Bin 16.

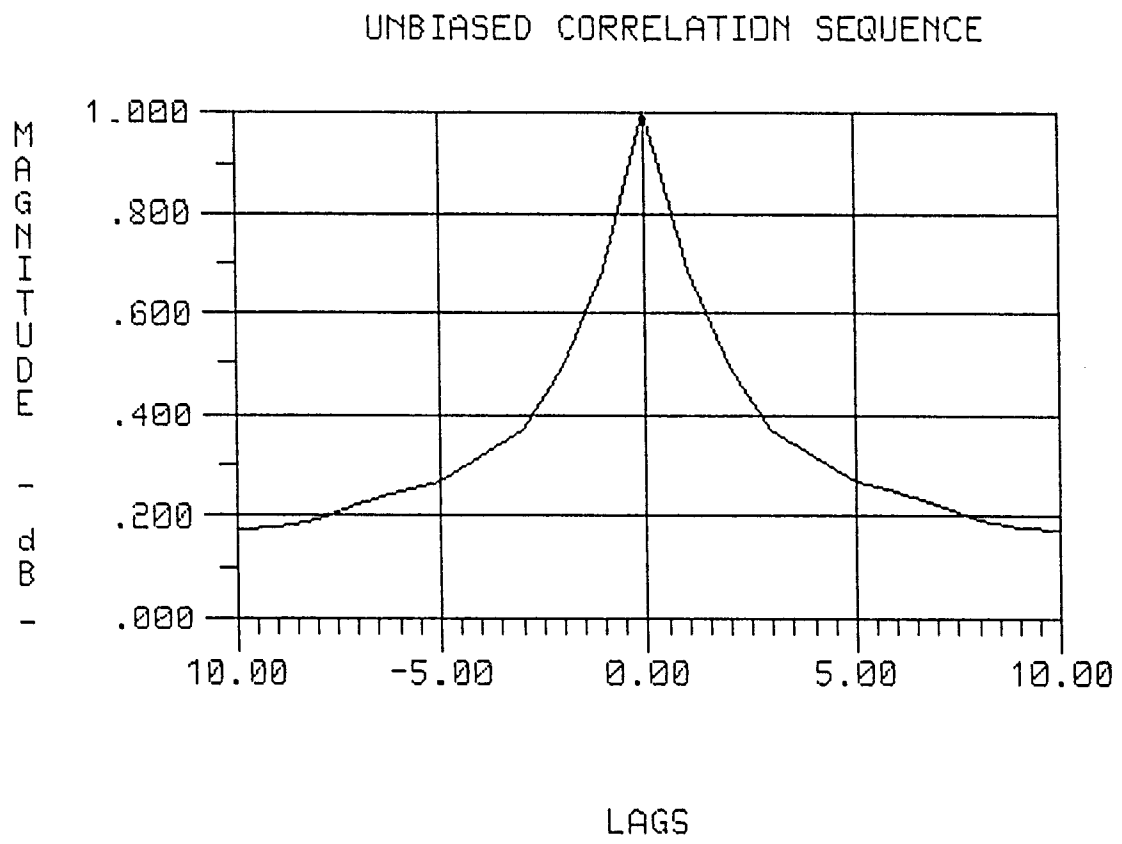


Figure 51. Correlation Sequence Estimate of Bin 16 Data.

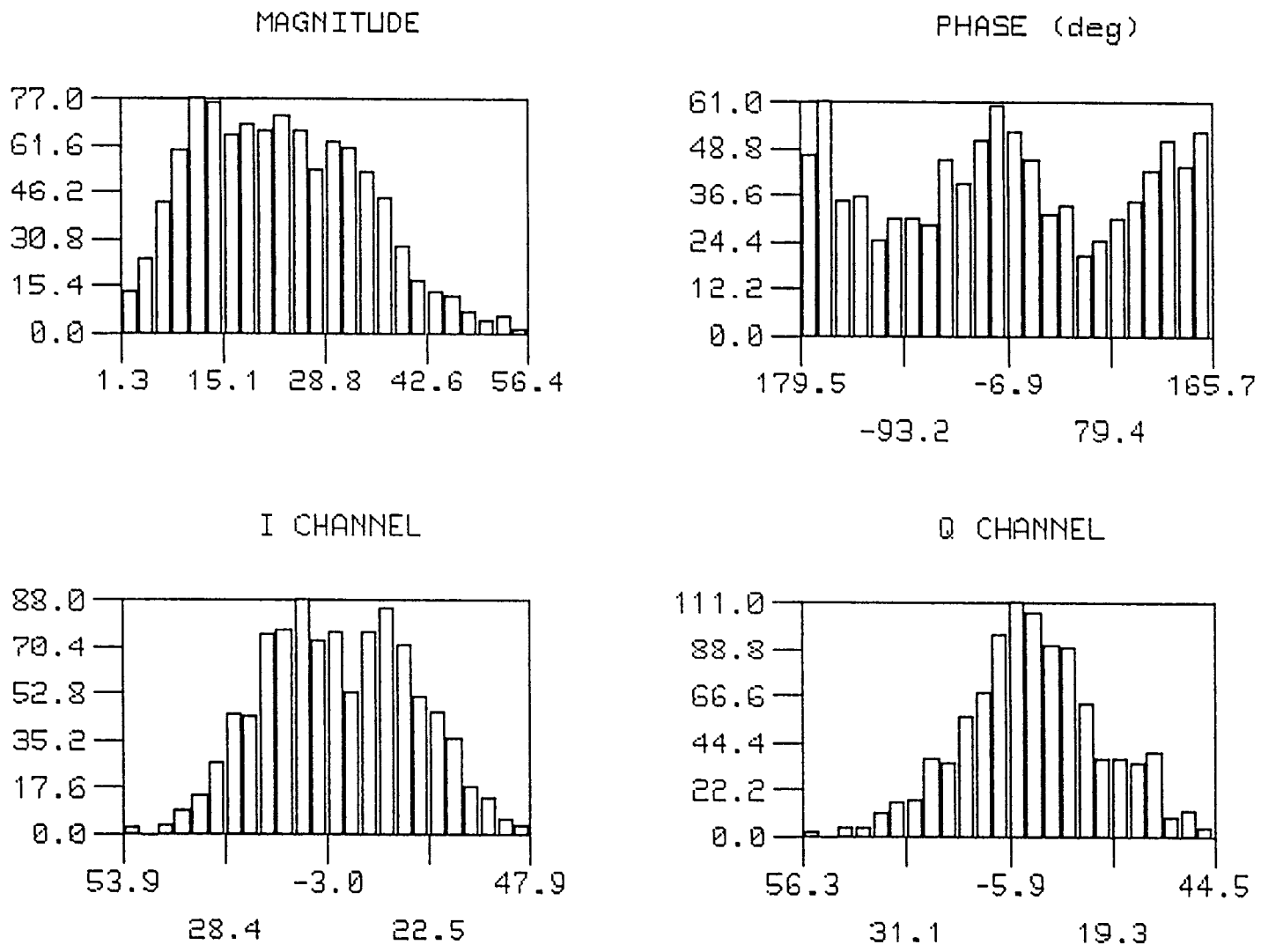


Figure 52. Histograms of Bin 16 Raw Data.

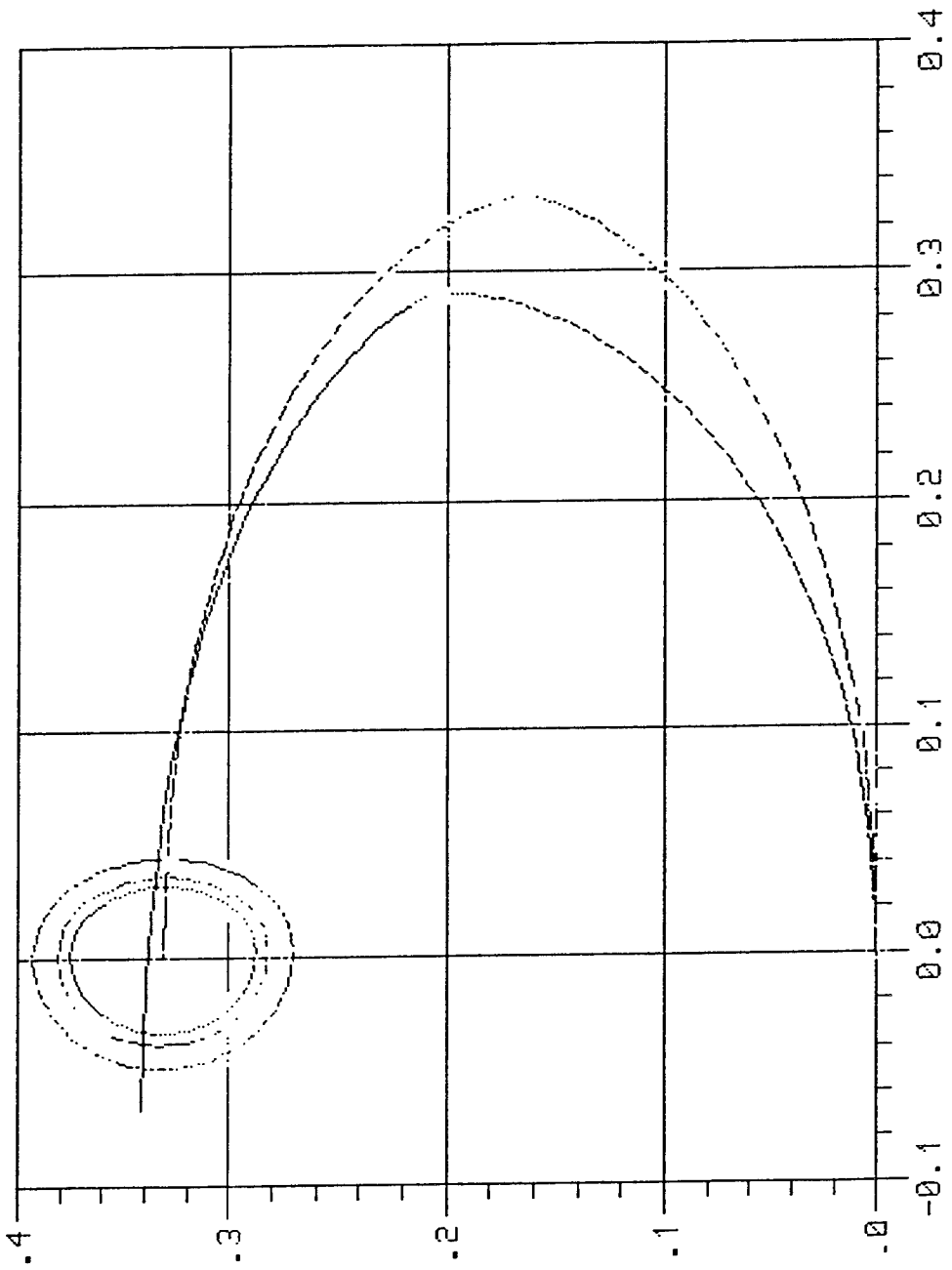
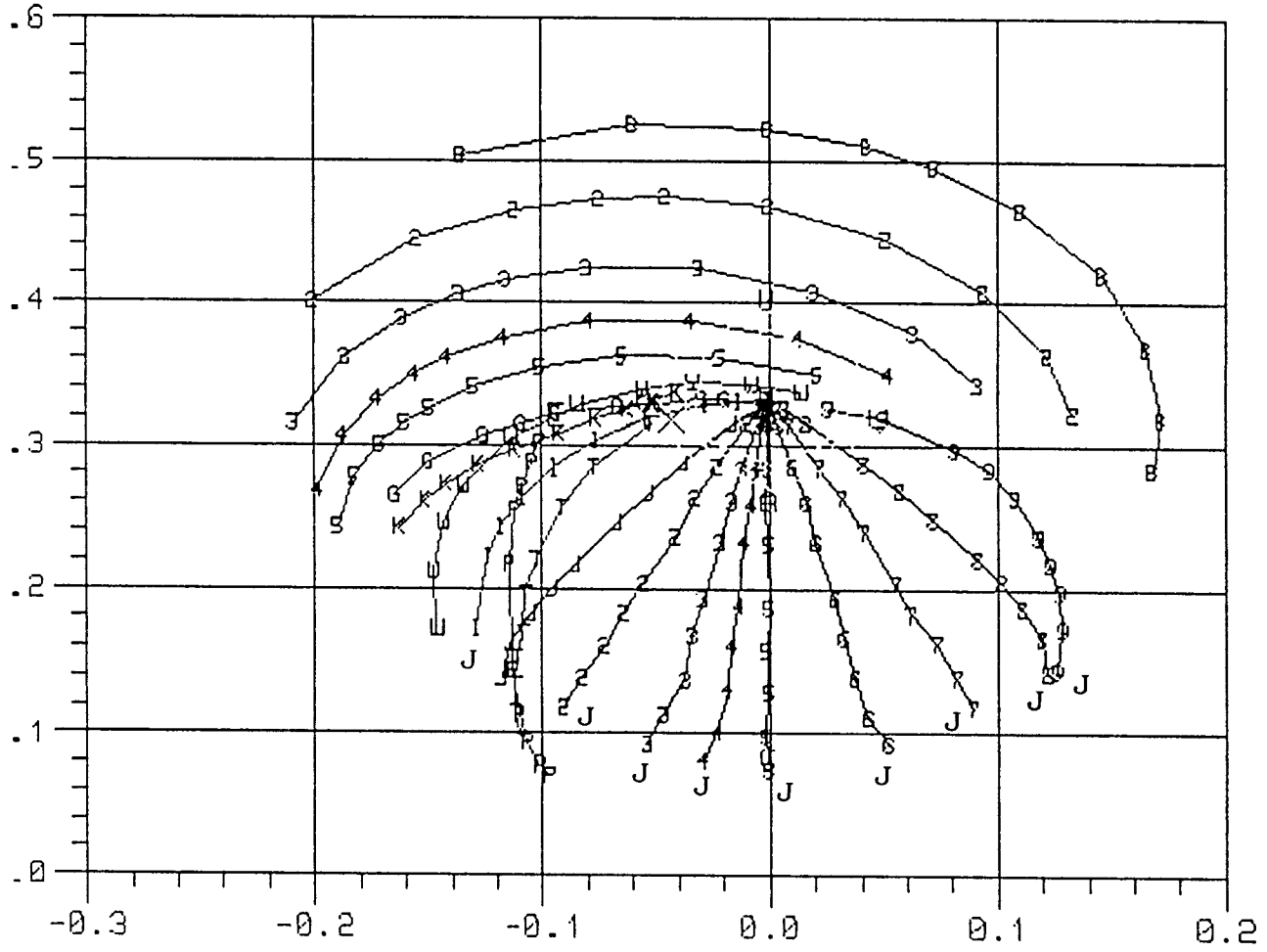


Figure 53. Goodness-Of-Fit Plot for Bin 16.



Nomenclature Indicates Type of Distribution

B, 2, 3, 4, and 5 – Beta Distribution with different values for the first shape parameter
G – Gamma Distribution
K – K Distribution
W – Weibull Distribution
P – Pareto Distribution
T – Extreme Value Distribution
L – LogNormal Distribution
U – Uniform Distribution
E – Exponential Distribution
J – SU-Johnson System (family of 9 curves)

Figure 54. PDF Approximation Chart for Bin 16.

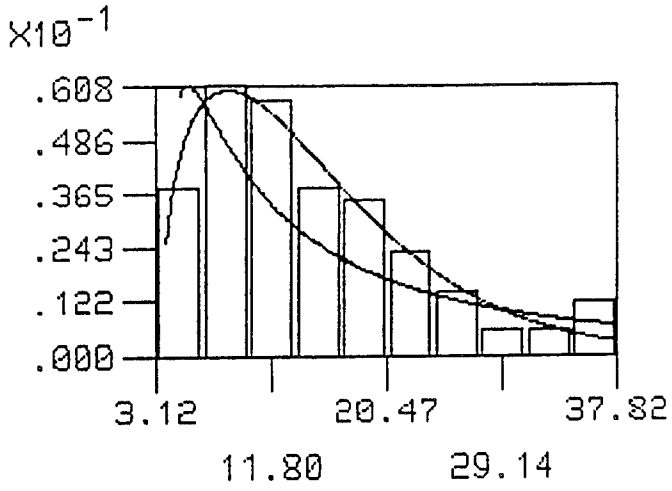


Figure 55. Overlay of Best/Worst PDF Approximations for Bin 16.

References

1. Papa, Robert J., Lennon, John F., and Taylor, Richard L. (1986) The Variation of Bistatic Rough Surface Scattering Cross Section for a Physical Optics Model, *IEEE Trans. Antennas Propagat.*, **AP-3**, (No. 10).

2. Sharpe, Lisa M. (1991) *Analytical Characterization of Bistatic Scattering From Gaussian Distributed Surfaces*, RL-TR-91-351, AD254253.

3. Shah, Rajiv R. (1993) *A New Technique for Distribution Approximation of Radar Data*, M.S. Thesis, Syracuse University.

4. Slaski, Lisa, and Rangaswamy, Muralidhar, (RL Report in Preparation) *An Introduction to Dr. Ozturk's Algorithm for PDF Approximation.*

5. Rangaswamy, M., Chakravarthi, P., Weiner, D.D., Cai, L., Wang, H., and Ozturk, A. (1993) *Signal Detection in Correlated Gaussian and Non-Gaussian Radar Clutter*, RL-TR-93-79, AD267453.

Appendix

A New Method for Univariate Distribution Approximation

A1. INTRODUCTION

In this appendix we address the problem of approximating the PDF of a set of random data. In practice, the clutter PDF encountered in radar signal processing is not known *a priori*. Consequently, a scheme that approximates the clutter PDF based on a set of measured data is necessary. Currently, available tests such as the Kolmogorov-Smirnov test and the Chi-Square test address the problem of goodness-of-fit for random data. In particular, these tests provide information about whether a set of random data is statistically consistent with a specified distribution, to within a certain confidence level. However, if the specified distribution is rejected, these tests cannot be used for approximating the underlying PDF of the random data. Moreover, these tests require large sample sizes for reliable results.

In practice, only a small number of samples may be available. Therefore, the scheme used should be efficient for small sample sizes. Ozturk has developed a new algorithm based on sample order statistics[1] for univariate distribution identification. This algorithm has two modes of operation. In the first mode the algorithm performs a goodness-of-fit test. Specifically, the test determines, to a desired confidence level, whether random data

[1]Ozturk, A., A new method for univariate and multivariate distribution identification, Submitted for publication to *J. Amer. Statistical Assn.*

are statistically consistent with a specified probability distribution. In the second mode of operation the algorithm approximates the PDF underlying the random data. By analyzing the random data and without any *a priori* knowledge, the algorithm identifies, from a stored library of PDFs, the particular density function that best approximates the data. Estimates of the scale, location, and shape parameters of the PDF are provided by the algorithm. The algorithm typically works well with small sample sizes of between 50 and 100 samples. An extension of this algorithm for the multivariate Gaussian PDF has been considered in Ozturk,[1] and Ozturk and Romeu.[2]

In this appendix we describe a new method for univariate distribution approximation. In Section A2 we present definitions. Section A3 describes the algorithm developed by Ozturk for univariate distribution identification. The proposed distribution identification algorithm is discussed in Section A4. Section A5 proposes a method to estimate the shape parameter based on the procedure developed in Section A4. Finally, conclusions are presented in Section A6.

A2. DEFINITIONS

Let $f_Y(y)$ denote the PDF of Y which has been standardized in a specified manner. Introduce the linear transformation defined by

$$x = \beta y + \alpha \tag{A.1}$$

The PDF of X is given by

$$f_X(x) = \frac{1}{|\beta|} f_Y(\frac{x-\alpha}{\beta}) \tag{A.2}$$

where α and β are defined to be the location and scale parameters of X, respectively. The mean μ_x and variance σ_x of the random variable X are given by

$$\mu_x = E(X)$$
$$\sigma_x^2 = E[(X-\mu_x)^2] \tag{A.3}$$

where E denotes the expectation operator. Although the mean and the variance are related to the location and scale parameters, note that the location parameter is not the mean value and the scale parameter is not the square root of the variance, in general. However, for a standardized Gaussian PDF $f_Y(y)$ for which the mean is zero and the variance is unity, the location parameter is the mean of X and the scale parameter is the standard deviation (square root of the variance) of X.

[2]Ozturk, A. and Romeu, J., A new method for assessing multivariate normality with graphical applications, Accepted for Publication in *Commun. in Statistics*.

The coefficient of skewness, α_3, and the coefficient of kurtosis, α_4, are defined to be

$$\alpha_3 = \frac{E[(X-\mu_x)^3]}{\sigma_x^3}$$

$$\alpha_4 = \frac{E[(X-\mu_x)^4]}{\sigma_x^4}. \tag{A.4}$$

It is readily shown that α_3 and α_4 are invariant to the values of μ_x and σ_x. For any PDF that is symmetric about the mean, $\alpha_3 = 0$. For the Gaussian distribution, $\alpha_3 = 0$ and $\alpha_4 = 3$.

A3. GOODNESS OF FIT TEST

In this section, we introduce a general graphical method for testing whether a set of random data is statistically consistent with a specified univariate distribution. The proposed method not only yields a formal goodness-of-fit test but also provides a graphical representation that gives insight into how well the random data is represented by the specified distribution (null hypothesis). Using the normal distribution as *a reference distribution*, the standardized sample order statistics are represented by a system of linked vectors. Both the terminal point of these linked vectors and the shape of their trajectories are used in determining whether or not to accept the null hypothesis.

In this section we first give a brief description of the corresponding test statistic and then explain the goodness of fit test procedure. For illustration purposes, we assume that the null distribution is Gaussian. However, the proposed procedure works for any null hypothesis.

Let X_k; $k = 1, 2, \ldots n$ denote the k^{th} sample from a Gaussian distribution with mean μ and variance σ^2. We define

$$Y_k = \frac{X_k - \overline{X}}{S} \quad k = 1, 2, \ldots, n \tag{A.5}$$

where $\overline{X} = \Sigma X_k/n$ is the sample mean and $S = \{\Sigma(X_i - \overline{X})^2/(n-1)\}^{1/2}$ is the sample standard deviation. The standardized order statistics are denoted by $Y_{i:n}$ $i = 1, 2, \ldots n$ and are obtained by putting the Y_k; $k = 1, 2, \ldots n$ in a monotonic nondecreasing order so that $Y_{1:n} \leq Y_{2:n} \leq \ldots \leq Y_{n:n}$. This sequence is called the order statistics of $Y_1, Y_2, \ldots Y_n$. $Y_{i:n}$ is called the i^{th} order statistic. The i^{th} linked vector is characterized by its length and orientation with respect to the horizontal axis. Let $X_{1:n} \leq X_{2:n} \leq \ldots \leq X_{n:n}$ denote the ordered samples obtained by ordering X_k; $k = 1, 2, \ldots n$. Let $m_{1:n}, m_{2:n}, \ldots, m_{n:n}$ denote the expected values of the standard normal order statistics, where $m_{i:n} = E[\frac{(X_{i:n}-\mu)}{\sigma}]$. The length of the i^{th} vector a_i is obtained from the absolute value of the i^{th} standardized sample order statistic $Y_{i:n}$, while its orientation θ_i is related to $m_{i:n}$. By definition,

$$a_i = \frac{|Y_{i:n}|}{n}$$

$$\theta_i = \pi\Phi(m_{i:n}) \tag{A.6}$$

where $\Phi(x) = (\sqrt{2\pi})^{-1} \int_{-\infty}^{x} exp(-\frac{t^2}{2})dt$ is the distribution function of the standard Gaussian distribution. We define the sample points in a two dimensional plane by

$$Q_k = (U_k, V_k) \quad k = 1, 2, \ldots n \tag{A.7}$$

where $U_0 = V_0 = 0$ and

$$U_k = \frac{1}{k}\Sigma_{i=1}^{k}\{Cos(\theta_i)\}|Y_{i:n}|$$

$$V_k = \frac{1}{k}\Sigma_{i=1}^{k}\{Sin(\theta_i)\}|Y_{i:n}| \tag{A.8}$$

k=1,2,... n.

The sample linked vectors are obtained by joining the points Q_k. Note that $Q_0 = (0, 0)$. It should also be noted that the statistic Q_n given in Eq. (A.7) represents the terminal point of the linked vectors defined above. Figure A1 shows the linked vectors obtained for the Gaussian distribution with $n = 6$. The null distribution was obtained by averaging the results for 50,000 Monte Carlo trials. The solid curve in Figure A.1 shows the linked vectors for the sample distribution while the dashed curve shows the linked vectors for the null distribution. The magnitudes and angles of the linked vectors are obtained from Eq. (A.6). Note that the angles are independent of the data and depend only on the sample size n. Only the magnitudes of the linked vectors depend on the samples drawn and change from one trial to another.

For a typical set of ordered samples (that is, ordered samples drawn from the null distribution) it is reasonable to expect that the sample linked vectors would closely follow the null pattern. If the ordered set of samples is not from the null distribution, the sample linked vectors are not expected to closely follow the null pattern. Hence, the procedure provides visual information about how well the ordered set of samples fit the null distribution.

An important property of the Q_n statistic is that it is invariant under linear transformation. In particular, we consider the standardization used in Eq. (A.5). Let $Z_i = aX_i + b$, where a and b are known constants. Let S' denote the sample standard deviation of the samples Z_i. Then, it is readily shown that $\frac{|X_i - \overline{X}|}{S} = \frac{|Z_i - \overline{Z}|}{S'}$. The invariance property follows as a consequence. The advantage of this property is that the PDF of $Q_n = (U_n, V_n)$ depends only on the sample size n and is unaffected by the location and scale parameters. Since it is difficult to determine the joint PDF of U_n and V_n analytically, it is necessary to obtain empirical results.

Assuming that the conditions of the central limit theorem are satisfied, the marginal PDFs of U_n and V_n can be approximated as Gaussian, in the limit of large n. In addition, it is assumed that the joint PDF of U_n and V_n is approximately bivariate Gaussian. Consequently, all that is needed to determine the bivariate PDF is the specification of $E(U_n)$, $E(V_n)$, $E(U_nV_n)$, $Var(U_n)$ and $Var(V_n)$. Drawing samples from the Gaussian distribution, it has been shown empirically in Reference [1] that for $3 \leq n \leq 100$

$$
\begin{aligned}
&E(U_n) = 0 \\
&E(V_n) = \mu_v \approx 0.326601 + \frac{0.412921}{n} \\
&E(U_nV_n) = 0 \\
&Var(U_n) = \sigma_u^2 \approx \frac{0.02123}{n} + \frac{0.01765}{n^2} \\
&Var(V_n) = \sigma_v^2 \approx \frac{0.04427}{n} - \frac{0.0951}{n^2}.
\end{aligned}
\tag{A.9}
$$

Since U_n and V_n are approximately bivariate Gaussian for large or moderate sample sizes, their joint PDF can be written as

$$f_{U_n,\, V_n}(u_n,\, v_n) = (2\pi)^{-1}(\sigma_u\sigma_v)^{-1}exp(-\frac{t}{2}) \tag{A.10}$$

where

$$t = \frac{u_n^2}{\sigma_u^2} + \frac{(v_n - \mu_v)^2}{\sigma_v^2}. \tag{A.11}$$

Let $t = t_0$. Then the equation

$$t_0 = \frac{u_n^2}{\sigma_u^2} + \frac{(v_n - \mu_v)^2}{\sigma_v^2} \tag{A.12}$$

is that of an ellipse in the u_n, v_n plane for which

$$f_{U_n,\, V_n}(u_n,\, v_n) = (2\pi)^{-1}(\sigma_u\sigma_v)^{-1}exp(-\frac{t_0}{2}). \tag{A.13}$$

Points that fall within the ellipse correspond to those points in the u_n, v_n plane for which

$$\mathrm{f}_{U_n,\, V_n}(u_n,\, v_n) > (2\pi)^{-1}(\sigma_u\sigma_v)^{-1}exp(-\frac{t_0}{2}). \tag{A.14}$$

Let

$$\alpha = P(T > t_0) = P(u_n, v_n \textit{ both falling outside the ellipse given by Eq. (A.12)}). \quad \text{(A.15)}$$

It is well known that the PDF of the random variable T defined by Eq. (A.11) has a Chi-Square distribution with two degrees of freedom[3] and is given by

$$f_T(t) = 0.5exp(-\frac{t}{2}). \quad \text{(A.16)}$$

Hence,

$$\alpha = 1 - exp(-\frac{t_0}{2}). \quad \text{(A.17)}$$

Consequently, $t_0 = -2ln(1-\alpha)$. Thus, Eq. (A.12) becomes

$$\frac{u_n^2}{\sigma_u^2} + \frac{(v_n - \mu_v)^2}{\sigma_v^2} = -2ln(1-\alpha). \quad \text{(A.18)}$$

α is known as the significance level of the test. It is the probability that Q_n falls outside the ellipse specified by Eq. (A.18) given that the data come from a Gaussian distribution. $1-\alpha$ is known as the confidence level and the corresponding ellipse is known as the confidence ellipse.

Equation (A.12) can be written in the standardized form

$$1 = \frac{u_n^2}{\sigma_u^2 t_0} + \frac{(v_n - \mu_v)^2}{\sigma_v^2 t_0} \quad \text{(A.19)}$$

where the lengths of the major and minor axes are given by $max\ [\sigma_u\sqrt{t_0},\ \sigma_v\sqrt{t_0}]$ and $min\ [\sigma_u\sqrt{t_0},\ \sigma_v\sqrt{t_0}]$, respectively. From Eq. (A.17), observe that smaller values of α correspond to larger values of t_0. Consequently, the confidence ellipses become larger as the confidence level is increased.

For a given sample size n ($n \leq 100$) approximate values of μ_v, σ_u^2 and σ_v^2 can be obtained from Eq. (A.9). The confidence ellipse of Eq. (A.18) can then be used to make a visual test of the null hypothesis. If the terminal sample point falls inside the ellipse, then the data are declared consistent with the Gaussian distribution with confidence level $1-\alpha$. Otherwise the null hypothesis is rejected with a significance level α.

A major difficulty in determining the joint PDF of U_n and V_n is that the coefficients of skewness and kurtosis of U_n and V_n (see Table 1) indicate that the Gaussian approximation for the bivariate PDF may not be satisfactory for $n < 10$. The empirical bivariate PDF of U_n and V_n were obtained by using 50,000 Monte-Carlo trials for n=3, 10, 20, 30, 50 and 100. The corresponding probability contours are shown in Figure A2. The same procedure is used even when the null distribution is different from the Gaussian distribution.

[3]Johnson, N. and Kotz, S. (1976) *Distributions in Statistics: Continuous Multivariate Distributions*, New York: John Wiley and Sons.

However, note that the standard Gaussian distribution is always used as the reference distribution for determining the angles θ_i.

A4. DISTRIBUTION APPROXIMATION

In this section we present a graphical procedure for approximating the underlying PDF of a set of random data based on the goodness-of-fit test procedure discussed in Section A3.

Following a similar approach to that outlined in Section A3, random samples are generated from many different univariate probability distributions. For each specified distribution and for a given n, the statistic $Q_n = (U_n, V_n)$ given by Eq. (A.8) is obtained for various choices of the shape parameter. Thus, each distribution is represented by a trajectory in the two dimensional plane whose coordinates are U_n and V_n. Figure A3 shows an example of such a representation. Twelve distributions, namely Gaussian (1), Uniform (2), Exponential (3), Laplace (4), Logistic (5), Cauchy (6), Extreme Value (7), Gumbel type-2 (8), Gamma (9), Pareto (10), Weibull (11) and Lognormal (12), are represented in this chart. The value of Q_n at each point of the trajectories is obtained by Monte-Carlo experiments using the standard Gaussian distribution as the reference distribution for determining the angles θ_i. The results are based on averaging 1000 trials of 50 samples from each distribution. The samples from each distribution are obtained by using the IMSL subroutines for specified values of the shape parameter. Since the procedure is location and scale invariant, the trajectory reduces to a single point for those PDFs which do not have shape parameters but are characterized only in terms of their location and scale parameters. By way of example, the Gaussian, Laplace, Exponential, Uniform and Cauchy PDFs are represented by single points in the $U_n - V_n$ plane. However, those PDFs having shape parameters are represented by trajectories. For a given value of the shape parameter, a single point is obtained in the $U_n - V_n$ plane. By varying the shape parameter, isolated points are determined along the trajectory. The trajectory for the PDF is obtained by joining these points. In a sense the trajectory represents a family of PDFs having the same distribution but with different shape parameter values. For example, the trajectory corresponding to the Gamma distribution in Figure A3 is obtained by joining the points for which the shape parameters are 0.2, 0.3, 0.5, 0.7, 1.0, 2.0, 3.0, 4.0, 6.0, and 10.0. As the shape parameter increases, note that the Gamma distribution approaches the Gaussian distribution. The representation of Figure A3 is called an identification chart. Some distributions such as the β distribution and the SU-Johnson system of distributions, have two shape parameters. For these cases, the trajectories are obtained by holding one shape parameter fixed while the other is varied. For these distributions, several different trajectories are generated in order to cover as much of the $U_n - V_n$ plane as possible. For certain choices of the shape parameters, two or more PDFs become identical. When this occurs, their trajectories intersect on the identification chart.

The identification chart of Figure A3 provides a one to one graphical representation for each PDF for a given n. Therefore, every point in the identification chart corresponds to a specific distribution. Thus, if the null hypothesis in the goodness-of-fit test discussed in Section A3 is rejected, then the distribution that approximates the underlying PDF of the set of random data can be obtained by comparing Q_n obtained for the samples with the existing trajectories in the chart. The closest point or trajectory to the sample Q_n is chosen as an approximation to the PDF underlying the random data. The closest point or trajectory to the sample point is determined by projecting the sample point Q_n to neighboring points or trajectories on the chart and choosing that point or trajectory whose perpendicular distance from the sample point is the smallest. The complete approximation algorithm is summarized as follows.

1. Compute Y_k as specified in Section A3
2. Obtain the standardized order statistic $Y_{i:n}$.
3. Compute U_n and V_n from Eq. (A.8).
4. Obtain an identification chart based on the sample size n as discussed in this Section. Plot the sample point Q_n on this chart.
5. Compare the sample point Q_n with the existing distributions on the chart. The nearest neighboring point (or trajectory) on the chart is used as an approximation to the PDF of the samples.

The accuracy of this procedure can be increased by including as many distributions as possible in the identification chart. However, it is emphasized that this procedure does not identify the underlying PDF. Rather it identifies a suitable approximation to the underlying PDF.

A5. PARAMETER ESTIMATION

Once the distribution of the samples is approximated, the next step is to estimate its parameters. The method discussed in Section A4 lends itself for estimating the parameters of the approximated distribution. We present the estimation procedure for the location, scale, and shape parameters in this section.

A5.1 Estimation of Location and Scale Parameters

Let $f(x;\alpha,\beta,)$ denote the distribution which approximates the PDF of the set of random data, where α and β are the location parameter and scale parameter, respectively

of the approximating PDF. Let $X_{i:n}$ denote the ordered statistics of X from a sample of size n. The standardized ordered statistics are defined by

$$W_{i:n} = \frac{X_{i:n} - \alpha}{\beta}. \tag{A.20}$$

Let

$$\mu_{i:n} = E[W_{i:n}]. \tag{A.21}$$

Then

$$E[X_{i:n}] = \beta\mu_{i:n} + \alpha. \tag{A.22}$$

We consider the following statistics

$$\begin{aligned} T_1 &= \textstyle\sum_i Cos(\theta_i) X_{i:n} \\ T_2 &= \textstyle\sum_i Sin(\theta_i) X_{i:n} \end{aligned} \tag{A.23}$$

where θ_i is the angle defined in Eq. (A.6). The expected values of T_1 and T_2 are

$$\begin{aligned} E[T_1] &= \textstyle\sum_i Cos(\theta_i)[\beta\mu_{i:n} + \alpha] \\ E[T_2] &= \textstyle\sum_i Sin(\theta_i)[\beta\mu_{i:n} + \alpha]. \end{aligned} \tag{A.24}$$

These can be written as

$$\begin{aligned} E(T_1) &= a\alpha + b\beta \\ E(T_2) &= c\alpha + d\beta \end{aligned} \tag{A.25}$$

where

$$\begin{aligned} a &= \textstyle\sum_i Cos(\theta_i) \\ b &= \textstyle\sum_i \mu_{i:n} Cos(\theta_i) \\ c &= \textstyle\sum_i Sin(\theta_i) \\ d &= \textstyle\sum_i \mu_{i:n} Sin(\theta_i). \end{aligned} \tag{A.26}$$

Because the standardized Gaussian distribution is used as the reference distribution for θ_i, it can be shown that $a = 0$. It follows that

$$\hat{\beta} = \frac{E[T_1]}{b}$$
$$\hat{\alpha} = \frac{E[T_2 - d\hat{\beta}]}{c} \tag{A.27}$$

where the symbol $\wedge$ is used to denote an estimate. For n sufficiently large (that is, $n > 50$), suitable estimates for $E[T_1]$ and $E[T_2]$ are

$$\hat{E}[T_1] = T_1$$
$$\hat{E}[T_2] = T_2. \tag{A.28}$$

Estimates for b and d rely upon an estimate of $\mu_{i:n}$. We obtain $\hat{\mu}_{i:n}$ from a Monte Carlo simulation of $W_{i:n}$ where $W_{i:n}$ is generated from the known approximating distribution $f(x; 0, 1)$ having zero location and unity scale parameters. based upon 1000 Monte Carlo trials $\hat{\mu}_{i:n}$ is the sample mean of $W_{i:n}$ with $\hat{\mu}_{i:n}$ known, the estimates for b and d are given by

$$\hat{b} = \sum_i^n \hat{\mu}_{i:n} Cos(\theta_i)$$
$$\hat{d} = \sum_i^n \hat{\mu}_{i:n} Sin(\theta_i). \tag{A.29}$$

The scale and location parameters are then estimated by application of Eq. (A.27).

A5.2 Shape Parameter Estimation

In this section we present an approximate method for estimating the shape parameter of the approximating PDF. This procedure can be used only when one of the shape parameters is unknown. Let γ denote the shape parameter of the approximating PDF being estimated. Since U_n and V_n are location and scale invariant, the point Q_n depends only on the sample size n and the shape parameter γ. The expected value of U_n and V_n can be expressed as

$$E(U_n) = \varphi_1(n, \gamma)$$
$$E(V_n) = \varphi_2(n, \gamma) \tag{A.30}$$

where $\varphi_1(.,.)$ $\varphi_2(.,.)$ are some functions of γ and n. For a given sample size n and shape parameter γ_0 the corresponding expected point $\varphi_1(n, \gamma_0), \varphi_2(n, \gamma_0)$ can be determined approximately in the $U_n - V_n$ plane.

The proposed shape parameter estimation method is based on finding a point such that

$$U_n = \varphi_1(n, \hat{\gamma})$$
$$V_n = \varphi_2(n, \hat{\gamma}) \tag{A.31}$$

where $\hat{\gamma}$ is the sample estimator of γ. However, in many instances the sample point may not correspond exactly to a particular trajectory. In such a case, let $E(Q_{1n}) = (u_1, v_1)$ $E(Q_{2n}) = (u_2, v_2)$ denote the expected points corresponding to two different shape parameter values $\gamma = \gamma_1$ and $\gamma = \gamma_2$. It is assumed that the sample point lies in between the points corresponding to γ_1 and γ_2. Assuming that linear interpolation provides a satisfactory approximation, the estimate of the shape parameter corresponding to the sample point is given by

$$\hat{\gamma} \approx \gamma_1 + \frac{(\gamma_2 - \gamma_1)(x_0 - u_1)}{(u_2 - u_1)} \tag{A.32}$$

where

$$x_0 = \frac{\{A(V_n - v_1) + A^2 u_1 + U_n\}}{(A^2 + 1)}$$
$$A = \frac{(v_2 - v_1)}{(u_2 - u_1)}. \tag{A.33}$$

The accuracy of the procedure can be improved by employing a non-linear interpolation method. It must be emphasized that the shape parameter estimation procedure presented in this section is an approximate procedure.

A6. CONCLUSIONS

This appendix has presented a new algorithm for analyzing univariate random data. The algorithm provides a graphical goodness-of-fit test that determines whether a set of random data is statistically consistent with a specified PDF. Also, a graphical procedure is presented for approximating the underlying PDF of a set of random data. Estimation of location, scale and shape parameters of the approximating PDF have been discussed. Finally, it must be pointed out that the chief advantage of the algorithm presented in this appendix is that it works well for small sample sizes (between 50 and 100 samples).

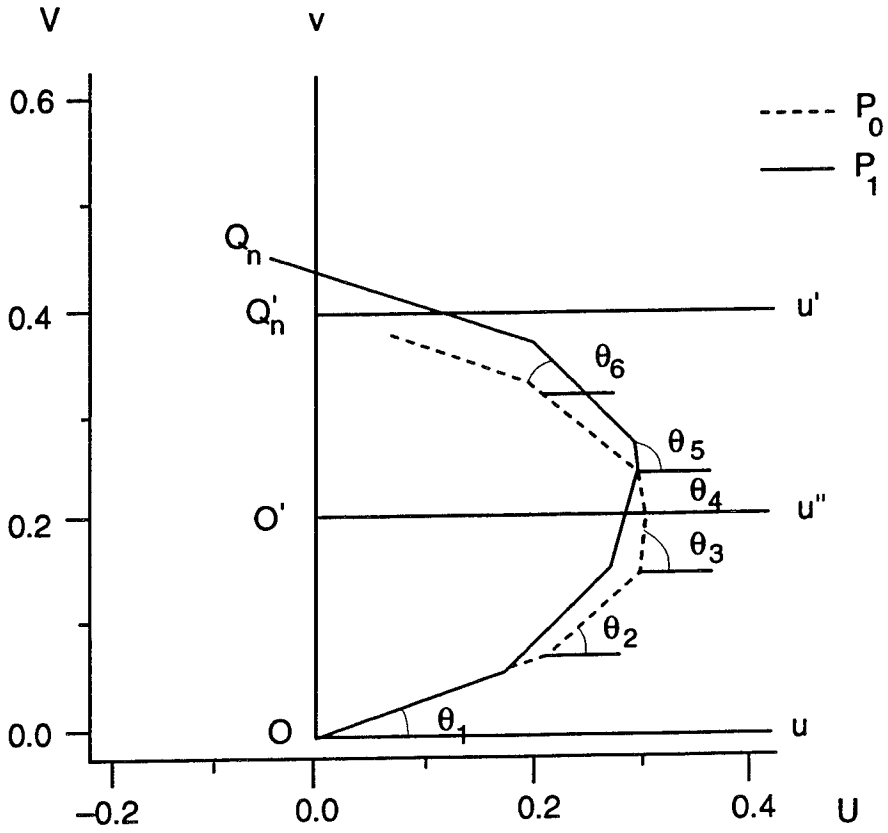


Figure A.1. Linked Vector Chart: Dashed Lines are for P_0 = Null Hypothesis Linked Vectors, for the Solid Lines P_1 = Linked Vectors for the Distribution being Sampled.

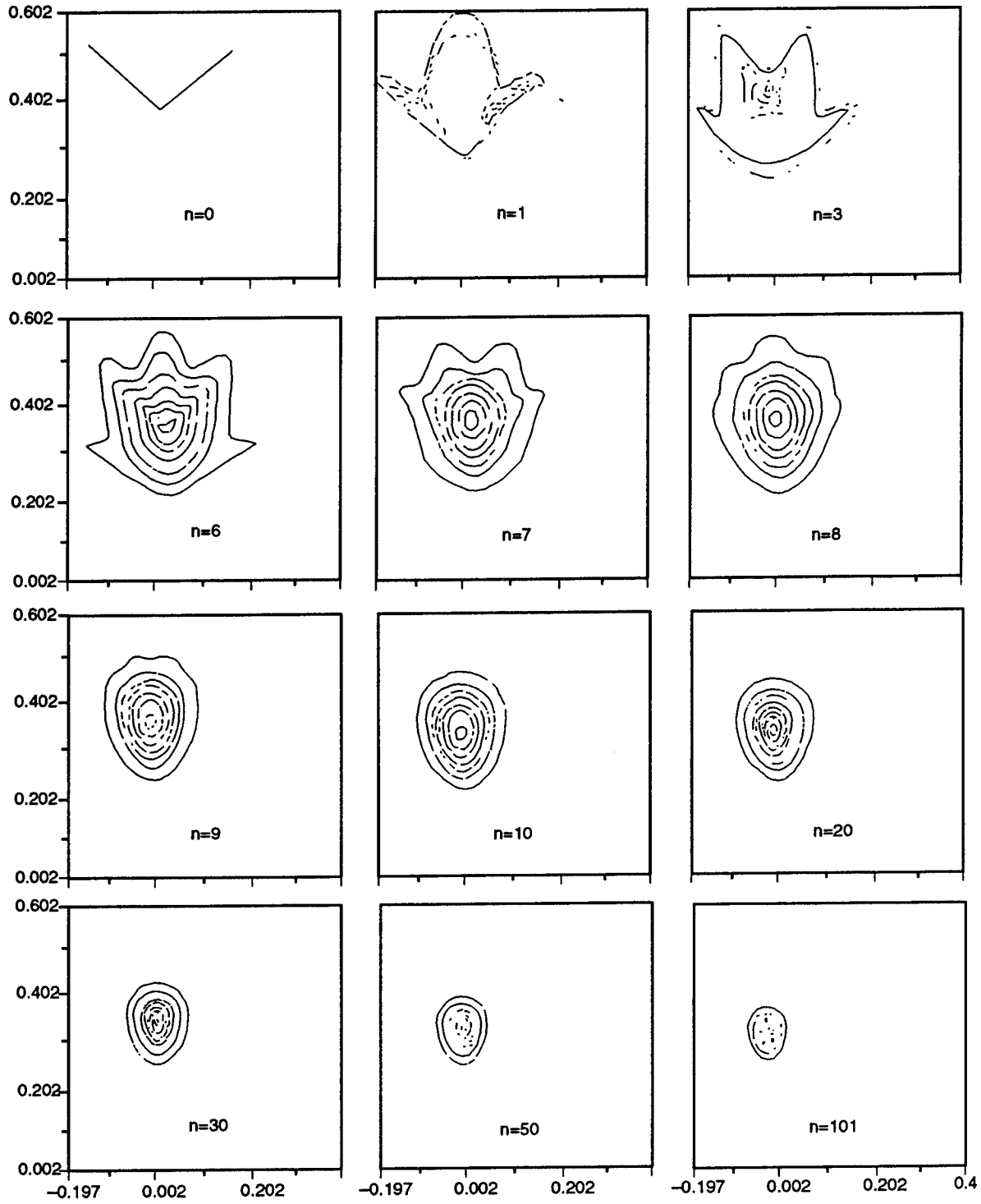


Figure A.2. Empirical Distribution of Q_n for Several Values of n.

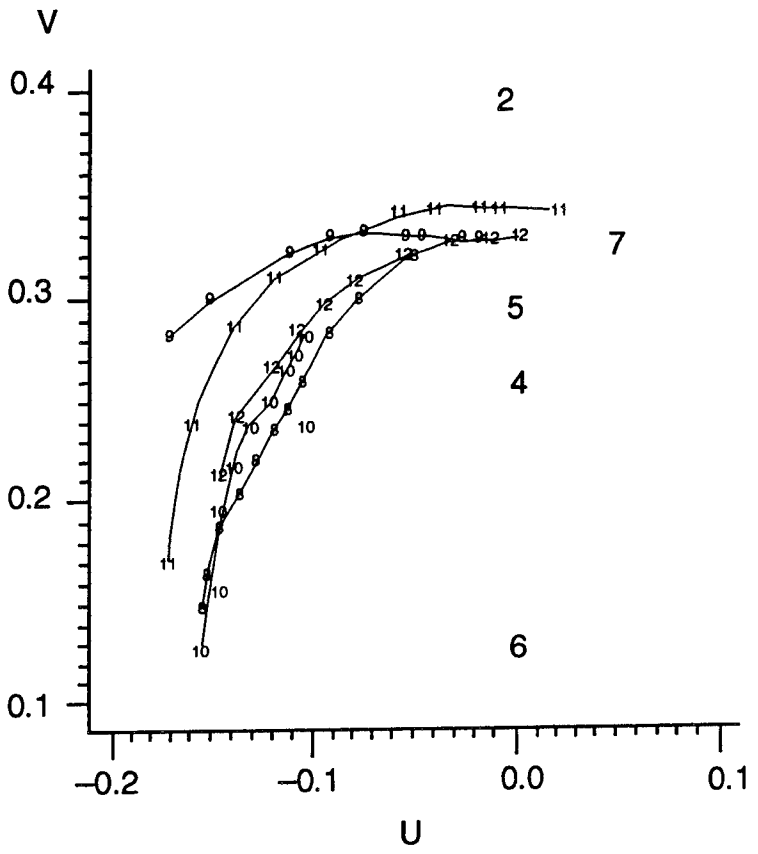


Figure A.3. Identification Chart for Univariate Distributions Based on 1000 Samples (n=50).

Appendix References

1. Ozturk, A., A new method for univariate and multivariate distribution identification, Submitted for publication to *J. Amer. Statistical Assn.*.

2. Ozturk, A. and Romeu, J., A new method for assessing multivariate normality with graphical applications, Accepted for Publication in *Commun. in Statistics*.

3. Johnson, N. and Kotz, S. (1976) *Distributions in Statistics: Continuous Multivariate Distributions*, New York: John Wiley and Sons.

MISSION

OF

ROME LABORATORY

Mission. The mission of Rome Laboratory is to advance the science and technologies of command, control, communications and intelligence and to transition them into systems to meet customer needs. To achieve this, Rome Lab:

a. Conducts vigorous research, development and test programs in all applicable technologies;

b. Transitions technology to current and future systems to improve operational capability, readiness, and supportability;

c. Provides a full range of technical support to Air Force Materiel Command product centers and other Air Force organizations;

d. Promotes transfer of technology to the private sector;

e. Maintains leading edge technological expertise in the areas of surveillance, communications, command and control, intelligence, reliability science, electro-magnetic technology, photonics, signal processing, and computational science.

The thrust areas of technical competence include: Surveillance, Communications, Command and Control, Intelligence, Signal Processing, Computer Science and Technology, Electromagnetic Technology, Photonics and Reliability Sciences.